Management methods 62:226

May 1933

Jon. P. [illegible]

A TEXT-BOOK

OF

PHYSICS

A TEXT-BOOK OF PHYSICS

A. WILMER DUFF, Editor

MECHANICS AND SOUND. By A. Wilmer Duff, D. Sc., Professor of Physics, Worcester Polytechnic Institute, Worcester, Massachusetts.

WAVE MOTION AND LIGHT. By E. Percival Lewis, Ph. D., Professor of Physics, University of California, Berkeley, California.

HEAT. By Charles E. Mendenhall, Ph. D., Professor of Physics, University of Wisconsin, Madison, Wisconsin.

ELECTRICITY AND MAGNETISM. By Albert P. Carman, D. Sc., Professor of Physics, University of Illinois, Urbana, Illinois.

CONDUCTION OF ELECTRICITY THROUGH GASES AND RADIO-ACTIVITY. By R. K. McClung, D. Sc., F. R. S. C., Assistant Professor of Physics, University of Manitoba, Winnipeg, Manitoba.

BLAKISTON'S SCIENCE SERIES

A TEXT-BOOK
OF
PHYSICS

EDITED BY

A. WILMER DUFF

CONTRIBUTORS

A. WILMER DUFF
ALBERT P. CARMAN
E. PERCIVAL LEWIS
R. K. McCLUNG
CHARLES E. MENDENHALL

FOURTH EDITION REVISED
WITH 609 ILLUSTRATIONS

PHILADELPHIA
P. BLAKISTON'S SON & CO.
1012 WALNUT STREET

THE MAPLE PRESS YORK PA

PREFACE TO THE FOURTH EDITION

In the present edition numerous changes have been made where in the experience of those who have used the book, the method of presentation could be simplified or clarified, especially in the paragraphs dealing with the Dynamics of Rotation. A new part on Sound has been prepared, attention being especially paid to recent important work by Professors Sabine and Miller and others. The editor wishes to express his indebtedness to Dr. Langmuir for information regarding his new air pump.

THE EDITOR.

WORCESTER, MASS.

EXTRACTS FROM PREFACE TO FIRST EDITION

The preparation of a work of this grade by the collaboration of several writers is a somewhat novel undertaking, and some explanation of its genesis will not be out of place. It represents the attempt of seven experienced teachers of college physics to prepare a text-book that would be more satisfactory to all of them than any existing one. It was, of course, hoped that such a book would also prove acceptable to other teachers. It seemed to the writers that there was a need, and there would be a place, for a work prepared in this way.

One or two remarks as to the character of the book may be permitted. It will in general be found that the writers, while aiming first of all at clearness and accuracy, have preferred terseness to diffuseness. Repetition and amplification are desirable in a lecture. In a printed statement, which may be reread and weighed until mastered, they often discourage thought; and a teacher of Physics might well begin his instruction with the words of Demosthenes, "In the name of the gods I beg you to think." The writers have endeavored to present their subjects simply and directly, avoiding, on the one hand, explanations obvious to any student of fair capacity, and, on the other hand, subtle distinctions and discussions suited to more advanced courses. Some may find the material included in the book too extensive for a single course. If so, a briefer course can be arranged by omitting all of the parts in small print together with as much of those in large print as may seem desirable. There may seem to be some duplication of topics in the work of two contributors. In such cases (which are very few), it will be found that the treatment is from different points of view, appropriate to the respective subdivisions of the subject.

THE EDITOR.

WORCESTER, MASS.

CONTENTS

MECHANICS AND PROPERTIES OF MATTER

WAVE MOTION

HEAT

ELECTRICITY AND MAGNETISM

LIST OF TABLES

Mechanics and Properties of Matter

Heat

Electricity and Magnetism

Radioactivity

Sound

Light

Greek Letters Used as Symbols

α	Alpha	θ	Theta	ρ	Rho
β	Beta	κ	Kappa	τ	Tau
γ	Gamma	λ	Lambda	ϕ	Phi
δ	Delta	μ	Mu	ω	Omega
η	Eta	π	Pi		

TEXT-BOOK OF PHYSICS

MECHANICS AND THE PROPERTIES OF MATTER

By A. Wilmer Duff, D. Sc.

Professor of Physics in the Worcester Polytechnic Institute, Worcester, Mass.

INTRODUCTION

1. Physics as a Science.—From the evidence of our senses we infer the existence of a great variety of bodies in the physical universe around us. By the use of our senses we also learn that these bodies have various characteristics in common, such as inertia, weight, and elasticity, and these we attribute to the *matter* of which in various forms all bodies seem to consist. Matter in itself is inert; the mutual actions of bodies and the effects which they produce on our senses are due to the presence in them of something which is not matter and which is called *energy*. We shall define the word energy later; the thing denoted by it is known to all as the means which are supplied by the sun, fuels, and elevated bodies of water, and which are required for various familiar operations in nature and industry.

Physics is the Science of the Properties of Matter and Energy.—This general description of Physics does not sharply distinguish it from Chemistry and, in fact, no definite dividing line can be drawn between the two sciences, although, in a general way, it may be stated that chemistry deals chiefly with questions regarding the composition and decomposition of substances. The different branches of Engineering also treat of the properties of matter, but from the point of view of their useful applications.

A science is more than a large amount of imformation on some subject. In very early times men must have had much valuable

information regarding the physical results of various actions and processes; but it was only when attempts were made to systematize and arrange this knowledge and to seek the relations between the different facts that the science of Physics began. The description of the phenomena of the physical world became more and more scientific as more numerous connections between physical phenomena were discovered and described. At the present time Physics has progressed farther in this direction than any other science, and, in seeking to give a brief account of the present state of the science of Physics, it must be our aim, not only to state the most important observed facts, but also to show the relations and interdependence of these facts.

It will be seen as we proceed that in some parts of the subject the relations between observed facts are better understood than in other parts. Thus in Mechanics the relations between phenomena have been so well ascertained that we are able to start from a few simple laws regarding the motions of bodies and from these deduce explanations of the most complicated motions. In other parts of the subject we must be content to take from time to time some one principle and trace the logical consequences of it as far as we can, and then proceed to do the same with other principles.

After classifying and studying a group of facts, the process by which we arrive at some underlying principle is called **Induction.** Thus, the principle of gravitation was discovered by Newton after a careful comparison of the motions of falling bodies and of the moon and the planets. Having found a general principle underlying and binding together many phenomena, we may reason forward from it and deduce other known or unknown facts, as in Geometry we deduce one proposition from another. This process is called **Deduction.** In a brief account of Physics we must necessarily use deductive more frequently than inductive methods; but, where space will permit, the effort may be made to show how by induction important fundamental principles have been discovered.

2. Measurement.—The first condition for success in tracing the connection between the facts in any science is that these facts shall be ascertained as accurately as possible. A qualitative statement of the size or weight of a body, to the effect that it is

large or small, is of very little use. A quantitative description of the same consists in giving the ratio of its size or weight to that of some accepted standard. Such a standard is called a unit, and the numerical ratio of the thing measured to the unit is called the numerical measure (or numeric) of the thing measured.

Some measurements are *direct*, that is, they are made by comparing the quantity to be measured directly with the unit of that kind, as when we find the length of a rod by placing a yard or meter scale beside it. But most measurements are *indirect*. For example, to measure the velocity of a train we measure the distance it travels and the time required, and by calculation we find the number of units of velocity in the velocity of the train.

3. Observation and Experiment.—In some branches of science mere *observation*, that is, taking note of circumstances and events, is the chief or only way of obtaining knowledge. For example, the astronomer cannot modify the motions of the heavenly bodies; he must be content to observe. Observation also plays an important part in Physics, but *experiment*, which consists in modifying circumstances or events with a view to making more valuable observations, plays a more important part. Thus, if we desire to know how the earth attracts a body and whether the attraction is different at different places, we cannot make much progress if we must confine ourselves to observing bodies falling freely from various heights; but, if we modify the fall by attaching the body to a cord and swinging it as a pendulum, we are able to make much more accurate observations, and to arrive at valuable information that we could probably never gain by observing free falling bodies. For this reason Physics is chiefly an experimental science, that is to say, the physicist relies on carefully planned experiments to find information and then, by methods of reasoning, and especially the condensed accurate form of reasoning called Mathematics, he extracts from the results of the experiment all the information possible.

4. Hypotheses.—An event or phenomenon remains obscure or unexplained when its logical connection with other events or phenomena has not been traced. But it is *explained* when it is shown to be connected with other familiar phenomena, and the nature of the connection is made clear. Thus, the rising of

mercury in an exhausted tube was obscure and unexplained until it was found to be different at different heights along a mountain side and to be connected with the pressure of the air on the mercury in the pool in which the tube stands. The explanation in such a case consists in tracing out the relation of cause and effect between the thing explained and other things. The latter may themselves be still unexplained. Thus the way in which air exercises pressure has only comparatively recently been explained.

A suggested explanation, while its correctness is still in doubt, is called an **hypothesis.** The hypothesis suggested to account for the pressure of air (or any gas) is that air consists of flying particles, which, by their bombardment of a surface, produce what we call the pressure on the surface; this suggested explanation is called the *kinetic hypothesis* of gases. The formation of an hypothesis plays a very important part in science, for it stimulates research to test its truth; and, even if this particular hypothesis turn out inadequate, in testing it many new facts are usually ascertained and the way is paved for arriving at the right explanation. The word **theory** is sometimes used in the same sense as hypothesis, but it is better to restrict it to meaning the extended discussion of an explanation or verified hypothesis. We shall use it in this sense later when speaking of the Kinetic Theory of Gases (§227).

5. Cause and Effect.[1]—When a certain event seems inevitably to be followed by a certain other event we are accustomed, in ordinary language, to speak of the former as the **cause** of the latter, and of the latter as the **effect** of the former. Thus, the explosion of powder in a gun is spoken of as the cause of the projection of the bullet, and the latter event is described as the effect of the explosion. In speaking of the relation of two things as that of cause and effect, we do not merely mean that one has always been observed to follow the other, but we suppose that there is something invariable in the connection between them, that is, we imply our belief that nature will always act in the same way when the circumstances are the same. The principle

[1] There is here no attempt to use terms in a critical philosophical sense. The use of such words cannot be avoided in an elementary work without confusing circumlocution and they must be used here in their ordinary sense.

thus stated is often called that of the Uniformity of Nature. There are, however, two circumstances which must be considered as of no importance as regards the connection between causes and effects. These are *time* and *place*. The time of an event is, of course, never repeated, and nothing, so far as we know, ever comes again to exactly the same place, since the sun and all the planets are moving rapidly through space.

6. Physical Laws.—A careful study of any phenomenon usually enables us to state in a general way what will happen in certain circumstances. Very ancient observation led to the conclusion that bodies when unsupported fall toward the earth. Such a generalization is a *physical law*. A still wider study usually leads to a more general law. Thus, the study of falling bodies and of the motion of the moon and of the planets led Newton to the conclusion that each of two bodies is attracted toward the other. The aim of physical research is to obtain physical laws of increasing width and generality. Any such law is very imperfect until it can be stated in exact mathematical form, and this requires careful measurement. By measurement and calculation Newton arrived at the law of attraction between bodies called the Law of Universal Gravitation. Thus a physical law is simply a statement that, given a certain set of circumstances, certain events will follow or it is a statement of some aspect of the Uniformity of Nature.

7. Subdivisions of Physics. The Science of Physics may, for convenience, be divided into the following parts:

1. Mechanics.	**3. Heat.**	**6. Sound.**
2. Wave Motion.	**4. Electricity and Magnetism.**	**7. Light.**
	5. Radioactivity.	

The subject-matter of each of these parts will be described when that part is taken up.

MECHANICS

8. Mechanics is the branch of Physics which treats of the motions of bodies and the causes of changes in these motions. It is divided into two parts, one, called **Kinematics**, in which the various kinds of motion are described and studied, and the other,

called **Dynamics**, in which the causes of change of motion are studied. Kinematics, or the study of motion, differs from Geometry in having to consider the element of time. Dynamics is usually divided into two parts, **Kinetics** and **Statics**, the former dealing with bodies in motion and the latter with bodies which, though acted on by causes that tend to produce motion, remain at rest, owing to the fact that these influences counteract each other. (Some authors use the term Dynamics in the sense here assigned to Kinetics.) In the following elementary treatment of Mechanics it will not be convenient to treat the various parts of the subject quite separately; each will be taken up in turn as convenience and simplicity may seem to dictate.

KINEMATICS

The Geometry of Displacements

9. **Translation and Rotation.**—Motions may be divided into two kinds. A moving body has a motion of **translation** when every straight line in the body remains parallel to its original position. Thus, a train moving on a straight track and a sled moving down a uniform incline have motions of translation. In such a case all points in the body move in exactly the same way. Hence the motion of the body is completely described when the motion of any point in the body is given, and we may, therefore, in describing the motion of the body, treat it as a single particle located at a point.

A body has a motion of **rotation** when all points in the body travel in circles the centers of which lie in a straight line; the line is called the axis of rotation. This is the motion of a grindstone, a flywheel, or a swing. Any two points in such a body are at any moment moving differently (unless they lie in a plane through the axis and are equidistant from the axis); points farther from the axis move in larger circles and more rapidly than those nearer to the axis.

Many forms of motion are highly complex, but they may in all cases be considered as made up of translations and rotations.

Since the motion of a body which has translation without rotation is the same as that of a point, it is convenient to begin with a study of the motion of a point.

10. Position of a Point.—The position of a point is fixed by its distances, or distances and directions, from other points, lines, or surfaces. The simplest way of stating the position of a point is by giving its distance and direction from some other point which we may call the starting-point or *origin*.

When we confine our attention to points in a certain line, straight or curved, their positions may be assigned by giving the distance of each point from some assumed origin in that line. One direction away from the origin is taken as positive and the opposite direction as negative. For example, the position of any station on a railway line may be fixed by its distance, positive or negative, from some other station taken as origin.

When we confine our attention to points on a surface, plane or curved, the position of each point may be assigned by its distance and direction from some origin on the surface, or, what comes to the same thing, by its distance from each of two lines at right angles passing through the origin. For example, a point on the surface of the earth is described as being a certain distance east or west and a certain distance north or south from the origin.

For points not confined to any line or surface the position of each may be assigned by its distance and direction from some assumed origin in space, or, what comes to the same thing, its distances, positive or negative, from each of three planes intersecting at right angles at the origin.

In the first case position is assigned by one number, in the second by two and in the third by three. A point is said to have one **degree of freedom** when its motion is confined to a definite line, two degrees of freedom when it is confined to a definite surface and three degrees of freedom when it is not restricted in any way.

The above statements of position are statements of *relative position*, that is, statements of the relation of the position of a point to that of some other point taken as origin. Absolute position, or the position of a point without any reference, stated or implied, to any other point or framework of lines, could not be described and no definite meaning could be attached to it. In what follows the word position will always mean relative position, and, unless otherwise stated or implied, the point of reference will be some point on the surface of the earth.

11. Displacements.—A change of position is called a displacement. In describing a displacement we do not need to make any reference to the time in which the point moves from one position to the other. A description of a displacement consists in a statement of the *length* and *direction* of the straight line drawn from the first position of the point to its second position. Thus, when a point has moved from A to B, it has received a displacement, the magnitude of which is the length of the straight line AB and the direction of which is the direction of AB. This displacement we may denote by the symbol $\overrightarrow{AB}$ or $\overline{AB}$, the arrow or stroke being placed above AB to indicate that we are referring not merely to the length of the line AB, but also to its direction from A to B.

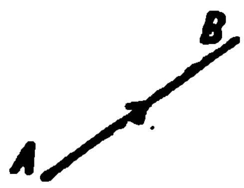


FIG. 1.—A displacement is represented by a directed line.

12. Units of Length. To measure or specify a displacement we must use some unit of length. The unit chiefly employed in Physics is the *meter* or one of its multiples or submultiples. The meter is defined as the distance between two lines on a bar of platinum-iridium kept at the International Bureau of Weights and Measures near Paris, when the temperature of the bar is that of melting ice. It was intended by the designers that this length should be one ten-millionth of the distance from a pole of the earth to the equator. One one-hundredth of the meter is called the centimeter (0.01 m.), and this is the unit of length which we shall usually employ. Other decimal fractions of the meter are the decimeter (0.1 m.) and the millimeter (0.001 m.). For great distances the kilometer (1000 m.) is employed.

The unit of length popularly used in English-speaking countries is the *yard* or one of its well-known multiples or submultiples. The British yard is defined legally as the distance between two lines on a bronze bar kept at the office of the Exchequer in London. The legal definition of the yard in the United States is $\frac{3600}{3937}$ of a meter (see Vol. I of the *Bulletin of the Bureau of Standards*, Washington, D. C.).

13. The Addition of Displacements.—If the point that moved from A to B did not travel by the straight line AB but passed through points C and D, its final displacement was the same as if

it had gone by the straight line AB; but the final displacement was the sum of a number of separate displacements, $\overline{AC}$, $\overline{CD}$, $\overline{DB}$. Thus $\overline{AB}$ is the *resultant* or sum of $\overline{AC}$, $\overline{CD}$, $\overline{DB}$, or we may say that by adding $\overline{AC}$, $\overline{CD}$, $\overline{DB}$ we get $\overline{AB}$, or briefly, $\overline{AB} = \overline{AC} + \overline{CD} + \overline{DB}$; but it must be carefully noted that the addition indicated by the sign + is a *geometrical* process, performed by placing the displacements end to end as the sides of a polygon and taking as the sum the displacement from the initial position to the final position.

If from C we draw a line CD' equal and parallel to DB, and from D' a line $D'B$ equal and parallel to CD, we shall have another path leading from A to B. The displacements $\overline{AC}$, $\overline{CD'}$, $\overline{D'B}$ added together give the same sum as the displacements $\overline{AC}$, $\overline{CD}$, $\overline{DB}$ added together, and for each step in one series there is an equal and parallel step in the other series. It is evident that, so far as addition of displacements is concerned, we may regard $\overline{CD'}$ and $\overline{DB}$ as the same displacement and $\overline{D'B}$ and $\overline{CD}$ as the same displacement. This is consistent with the definition of a displacement as a change of position; for, when a point goes from C to D, it has received the same *change* of position as another point has received when it has gone from D' to B, CD and $D'B$ being equal and parallel. Thus *all displacements which have the same magnitude and direction are equal.*

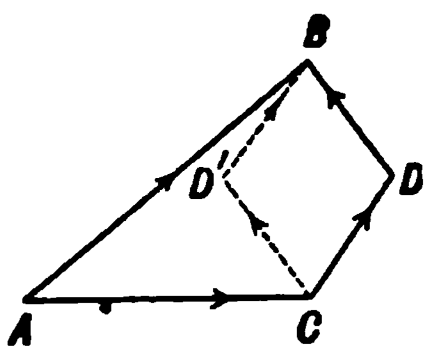


FIG. 2.—Geometrical addition of displacement.

When two displacements are to be added, the addition may be performed by drawing a triangle. Thus to add $\overline{AB}$ and $\overline{BC}$ we complete the triangle ABC and the sum is $\overline{AC}$. This is called the triangle method of adding two displacements. Another method of performing the addition is to construct a parallelogram. If AD be drawn from A equal and parallel to BC, the displacement $\overline{AD}$ is the same as the displacement $\overline{BC}$ and the sum of $\overline{AB}$ and $\overline{AD}$ is $\overline{AC}$, where AC is the diagonal of the parallelogram of which AB and AD are adjacent sides drawn away from A. This is called

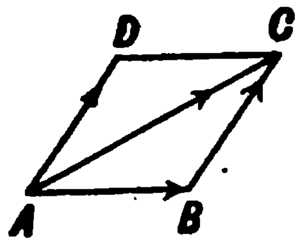


FIG. 3.

the parallelogram method of adding two displacements. When several displacements are to be added, the addition is performed by constructing a polygon as in Fig. 2.

14. Resolution and Subtraction of Displacements.—As we may replace any number of displacements by their geometrical sum or resultant, so we may replace a displacement by any number of displacements which added together give the original displacement. This is called *resolving a displacement into components.* Thus, to resolve a displacement $\overline{AC}$ (Fig. 3) into two components in given directions, we draw from A lines in the given direction and then complete the parallelogram $ABCD$ on the diagonal AC; $\overline{AB}$ and $\overline{AD}$ are the components desired, since their sum is $\overline{AC}$.

Subtraction is the opposite of addition. To subtract 4 from 10 we must find the number, 6, which added to 4 will give 10. Similarly, to subtract a displacement, $\overline{PQ}$, from another, $\overline{PR}$, we must find the displacement which added to $\overline{PQ}$ will give $\overline{PR}$. From the triangle method of addition this is evidently $\overline{QR}$, or, if we complete a parallelogram $PQRS$, it is $\overline{PS}$ which is equal to $\overline{QR}$. Denoting subtraction by the minus sign $\overline{PR}-\overline{PQ}=\overline{QR}=\overline{PS}$.

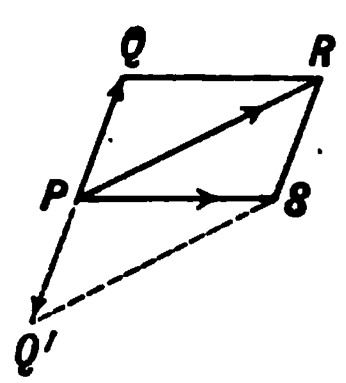


FIG. 4.—Subtraction of a displacement.

Subtraction may also be performed in a slightly different way. From the triangle method it is evident that $\overline{QP}$ added to $\overline{PR}$ will give QR. Hence to subtract a displacement we may *reverse its direction and add.* This addition may also be performed by a parallelogram $PQ'SR$ where $\overline{PQ'}=\overline{QP}$. Since subtracting $\overline{PQ}$ is the same thing as adding $\overline{QP}$, $-\overline{PQ}=+\overline{QP}$ or *the minus sign before a displacement reverses its direction.*

15. Vector Quantities and Vector Diagrams.—Displacements belong to the class of quantities called **vector quantities,** that is, quantities which have *magnitude* and *direction.* Other vector quantities are velocities, forces, etc. The figures in the preceding sections are diagrams of displacements, that is, they are made up of lines representing the actual displacements in magnitude and direction. Thus the diagram might be regarded as a reduced or enlarged

picture of the actual displacements. Other vector quantities, *e.g.*, a number of forces, may be similarly represented by a vector diagram by drawing lines each of which stands in magnitude and direction for one of the forces. The lines in such a diagram are called **vectors**. The lengths of any two vectors in such a diagram are to one another as the magnitudes of the forces represented, and the angle between the two vectors is the angle between these two forces. After we have defined the meaning of the resultant of a number of forces, it will be seen that it is represented as to magnitude and direction by the vector which is the sum of the vectors that represent the separate forces. Similar remarks apply to diagrams of velocities, accelerations, etc.

Quantities which imply no reference to direction are called **scalar** quantities. Such are mass, volume, etc. Each such quantity is assigned by a number without any idea of direction associated with it, and the addition or subtraction of such quantities is performed in the ordinary arithmetic or algebraic manner.

Velocity

16. Velocity is rate of change of position or *rate of displacement.* Since a displacement has a definite direction as well as a definite magnitude, a velocity also has a definite direction and a definite magnitude, or *velocities are vector quantities.* Thus "twenty miles an hour" is not a complete statement of a velocity, since it gives only the magnitude of the velocity and does not specify its direction; but "twenty miles an hour eastward" is a complete statement of a velocity. For clearness such a phrase as "twenty miles an hour" may be called the statement of a *speed,* which means the mere magnitude of a velocity or a rate of change of position without reference to the direction of the change. When the motions considered are all in the same straight line we do not need to distinguish speed and velocity.

17. Constant Velocity.—The velocity of a point is described as *constant* or *uniform* when the displacements of the point in all equal intervals of time are equal. By equal displacements must be understood displacements equal *in both magnitude and direction.* Hence, when the velocity of a point is constant, the point

moves in a straight line. The magnitude of a constant velocity is measured by the displacement in each unit of time. Hence, if we denote the magnitude of a constant velocity by v and the displacement in time t by s,

$$s = vt.$$

Unit velocity is the velocity of a point that travels unit distance in unit time, *e.g.*, 1 cm. in 1 sec.

18. Variable Velocity.—A point has a *variable velocity* when its displacements in equal times are not equal. The displacements in successive equal intervals of time may differ (1) in *magnitude only*, as when a point moves in a straight line with varying speed, or (2) in *direction only*, as when a point moves in a curve with constant speed, or (3) in *both magnitude and direction*, as when a point moves in a curve with varying speed. We shall begin by considering the first of these cases, that of rectilinear motion.

19. Average and Instantaneous Velocity.—In *rectilinear motion with variable velocity* how shall we define the magnitude of the velocity? In this case there are two ways open to us. If we divide the whole distance traversed in a certain interval of time by the length of the interval we get the *average velocity* in that interval. If for example we find the whole time required by a train to move from one station to another on a straight track and divide this into the whole distance, we get the *average velocity* between the two stations. In general, denoting the whole distance by s, the whole time by t, and the average velocity by $\bar{v}$, we have $\bar{v} = s/t$. Hence

$$s = \bar{v}t.$$

The magnitude of the average velocity in an interval tells us nothing as to the way in which the velocity varies during the interval. If we need to know the character of the motion more closely, we must divide the whole interval into parts and ascertain the average velocity in each. The smaller these parts, the more nearly does the average velocity in any one part represent the actual velocity at any moment in that part. Let us fix our attention on a certain moment at a time t after the beginning of the whole interval. If we proceeded to find the average velocity in a short interval, say Δt, including that moment, and if we took

successive decreasing values for Δt and found the average velocity in each of these decreasing values of Δt, we would find that the average velocity would rapidly approach a definite limiting value. This limiting value is the *instantaneous velocity* at the moment t. Stated more briefly, *if Δs is the displacement in a small interval of time Δt following the time t, the instantaneous velocity at the time t is the limiting value approached by $\Delta s/\Delta t$ as Δt approaches zero.* This may also be further abbreviated to the forms

$$v = \left[\frac{\Delta s}{\Delta t}\right]_{\Delta t = 0} = \frac{ds}{dt},$$

the last abbreviation being that used in the Differential Calculus.

When the velocity of a point is constant, the instantaneous velocity, as defined above, is the same as the velocity of the point, as defined in § 17. For the values of $\Delta s/\Delta t$ at different moments in any interval t are equal. Hence, if s is the whole distance traversed in the time t, each value of $\Delta s/\Delta t$ is equal to s/t, which is the distance traversed in unit time.

When the instantaneous velocity of a point is variable, we may also state its magnitude in terms of an equal constant velocity. Suppose that, when the instantaneous velocity is v, the point begins to move with a constant velocity equal to v. The magnitude of this constant velocity is the distance the point would travel in unit time. Hence we derive the statement that *the magnitude of the instantaneous velocity of a point is equal to the distance the point would travel in unit time if it had an equal constant velocity.*

20. The Unit of Time.—To measure or specify a velocity we must use some *unit of time.* The unit of time usually employed in Physics is the **mean solar second.** This is defined as $\frac{1}{86400}$ of the mean solar day, which is the average, thoughout a year, of the time between two successive transits of the sun across the meridian at any place. It is the second of the ordinary clock or watch when it is properly regulated.

21. Curvilinear Motion.—When the displacements of a point in successive equal intervals are in different directions, the point is moving in some curved path. This, for example, is the case when a ball is thrown obliquely upward or when a train is moving on a curved track. If the position of the point at a certain time t is

P and at a somewhat later time, say $(t+\Delta t)$, is Q, the displacement in this time is $\overline{PQ}$. If we denote the length of PQ by Δs and consider the limiting value of $\Delta s/\Delta t$ as before, we get the instantaneous velocity of the point at the time t when the point is at P. As PQ is decreased the chord PQ finally approaches without limit to the tangent at P; hence the direction of the instantaneous velocity at P is along the tangent at P. While this is the proper meaning of the rate of displacement at P, we should arrive at the same value for the instantaneous velocity if we took Δs to mean the length of the arc PQ, and supposed it successively diminished by the approach of Q toward P; for the chord and the arc would in the limit have a ratio of unity.

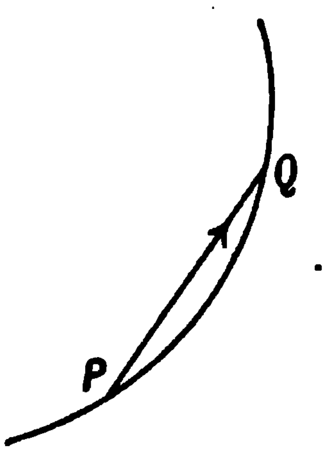


FIG. 5.

22. The Graph of the Speed of a Point.—When any quantity is variable, much valuable information can frequently be derived from the properties of a curve drawn to represent the varying quantity. A curve drawn to represent the speed of a moving point is called a *speed curve.* Let OA be a straight line of which the length OA stands for the length of the interval, t, in which we wish to consider the motion. Divide OA up into a very large number of small equal parts. At O erect a perpendicular OB to represent the speed at the beginning of the interval t. Erect similar perpendiculars to represent the instantaneous values of the speed at the beginnings of the other parts of the interval, and through the upper ends of these perpendiculars draw a smooth curve BC.

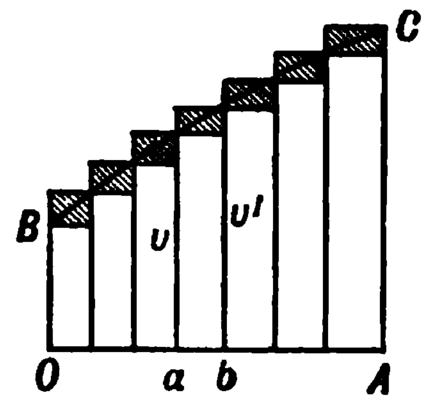


FIG. 6.—Graph of a speed.

Consider one of these short intervals, ab. If the speed throughout this short interval had been the same as at the beginning of the short interval, say v, the distance traversed in the short interval would have been $v \times ab$ or the unshaded rectangle above ab. If the speed throughout the short interval had been the same as that at the end of the short interval, say v', the distance would have been $v' \times ab$ or the area of the unshaded rectangle plus that of the small shaded rectangle above it. The real

distance in the interval is intermediate between these two. Applying the same reasoning to all the small intervals in succession, we see that the whole distance is something between that represented by the whole unshaded area between BC and OA and that represented by the unshaded area plus the shaded area. If the number of parts into which OA is divided be doubled, there will be twice as many small shaded rectangles, but the area of each will only be one-fourth as great as before. Hence if we suppose the number of the small intervals increased without limit, the shaded area will decrease without limit until it vanishes and the area between the curve BC and the line OA will represent the actual distance in the time t.

Since it is merely the magnitude of the velocity that is represented by each ordinate, the area represents the distance *measured along the line of motion*, whether this be straight or curved. Thus, if a point moves once around a circle with constant speed, BC will be a horizontal straight line, and the distance represented by the area $OBCA$ will equal the circumference; but the *mean velocity* in the revolution will be zero, since the final displacement will be zero.

To bring out more clearly the meaning to be attached to the word "represent" in the above, let us first suppose that OA contains as many units of length as t contains units of time, and that OB contains as many units of length as the velocity it stands for contains units of velocity. Each unit of area will then stand for a unit of distance traversed by the moving point, and the whole area will contain as many units of area as the distance traversed contains units of length. But if each unit of length along OA stands for m units of time and each unit of length along OB stands for n units of velocity, the whole area will be mn times smaller than it would have been on the first supposition, and, to get the whole distance, we shall have to multiply the whole area by mn.

23. The Resultant of Simultaneous Velocities.—A man sitting in a train has the velocity of the train, but, when he gets up and moves about, he has an additional velocity which may or may not be in the same direction as the first velocity. Similarly a launch floating down with the current in a river has the velocity of the current; but if it has a propeller in motion, it has another velocity in addition to the first. *When a body has two or more simul-*

taneous velocities, it pursues some definite path and its velocity in the path is called the **resultant** *of the simultaneous velocities.*

From this definition of the resultant of any number of simultaneous velocities it can be shown that the magnitude and direction of the resultant velocity can be deduced from the separate velocities by the triangle, parallelogram, or polygon method of adding vectors. Consider first the case of two constant velocities and draw a diagram, in which $\overline{AB}$ and $\overline{AC}$ stand for the two velocities. Complete the parallelogram $ABCD$. We shall show that $\overline{AD}$ stands for the resultant velocity. Since the velocities are constant $\overline{AB}$ and $\overline{AC}$ represent in magnitude and direction the component displacements in unit time, and the sum of these displacements is represented by $\overline{AD}$ (§ 13), which, therefore, represents *the resultant displacement in unit time.* Hence $\overline{AD}$ represents the resultant velocity. Thus the parallelogram method applies to the addition of constant velocities, and the same must be true of the other methods which are essentially the same.

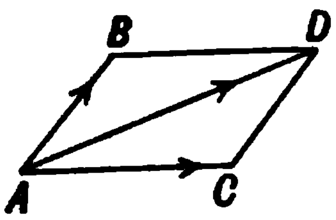


FIG. 7.

When the component velocities are not constant, we can add their instantaneous values by the vector methods referred to. The proof of this statement is the same as above, except that $\overline{AB}$, $\overline{AC}$, and $\overline{AD}$ will now stand for the displacements that would take place in unit time if the velocities remained constant that long.

24. Formula for Resultant.—Let v_1 and v_2 be the respective magnitudes of two component velocities of a moving point, and let these velocities be represented by $\overline{OA}$ and $\overline{OB}$ (Fig. 8). Also let v be the magnitude of the resultant velocity, which is represented by $\overline{OC}$, where OC is the diagonal of the parallelogram of which OA and OB are sides. By a well-known trigonometrical formula

$$OC^2 = OA^2 + AC^2 - 2OA \cdot AC \cos OAC$$

Denote the angle AOB, which is the angle between the directions of the two components, by θ. Then the angle OAC equals $(180° - \theta)$ and therefore $\cos OAC = -\cos \theta$. Since OA, OB, and OC are proportional to v_1, v_2, and v respectively,

$$v^2 = v_1{}^2 + v_2{}^2 + 2v_1 \cdot v_2 \cos \theta$$

By this formula we can calculate the magnitude of v when v_1, v_2, and θ are known.

When $\theta=0$, that is, when the components are in the same direction, $\cos\theta=1$ and the formula for v gives $v=(v_1+v_2)$. When $\theta=180°$, that is, when the components are in opposite directions, $\cos\theta=-1$ and $v=\pm(v_1-v_2$.

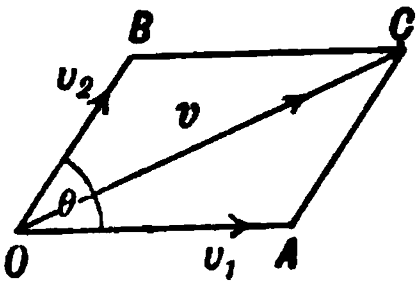


FIG. 8.

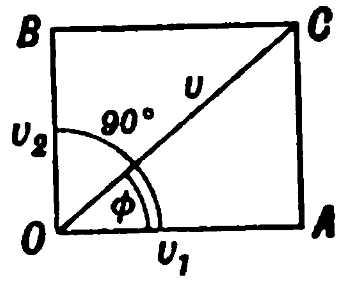


FIG. 9.

If $\theta=90°$, that is, if the components are at right angles, $\cos\theta=0$ and (Fig. 9)

$$v^2=v_1^2+v_2^2$$

and if ϕ be used to denote the angle AOC which the resultant makes with the component of magnitude v_1,

$$\tan\phi=\frac{AC}{OA}=\frac{v_2}{v_1}$$

25. Resolution of a Velocity into Components.—Since two velocities taken together are equivalent to a single velocity called their resultant, we may reverse the process and suppose any velocity replaced by any two velocities which added are equivalent to the original velocity. This is called *resolving a velocity into components*. To thus resolve a velocity we must draw a parallelogram of which the diagonal stands for the velocity to be resolved. Now any number of parallelograms can be drawn with a given line as diagonal; but, if the directions of the sides are specified, only one solution is possible. Hence to resolve a given velocity into components in two given directions is a definite problem, which may be solved graphically by constructing a parallelogram.

The most important case of the above is when the directions of the components are at right angles. Thus, if the velocity is v in the direction Oc and if $\overline{OC}$ is taken to represent v and if Oa and Ob are to be the directions of the components, we draw from C perpendiculars, CA and CB, to Oa and Ob respectively.

Then $\overline{OB}$ and $\overline{OA}$ are the desired components in the specified directions. If the direction Oa makes an angle θ with the direction of v and if we denote the components in the directions Oa and Ob respectively by v_1 and v_2,

$$v_1 = v \cos \theta. \quad v_2 = v \sin \theta.$$

It should be noted that θ stands for an angle that may be either positive or negative. We may regard θ as the angle through which a line, starting from the position Oa, must revolve about O to reach the position Oc; and when the revolution is counter-clockwise, it is customary to regard such an angle as positive, the opposite direction of revolution corresponding to a negative angle. If we make this agreement as regards the sign of θ, we must keep to it as regards the right angle that Ob makes with Oa, that is to say, the right angle and the angle θ must be measured away from Oa in the same direction, namely, counter-clockwise.

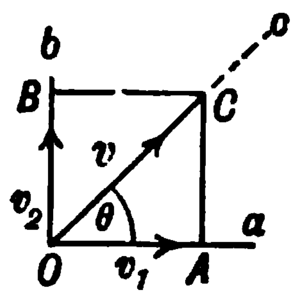


Fig. 10.

Acceleration

26. Acceleration is rate of change of velocity. A change of velocity has a definite direction as well as a definite magnitude. Hence acceleration is a quantity which has both direction and magnitude, that is, acceleration is a vector quantity.

An acceleration may be either constant or variable. The acceleration of a point is *constant* when the velocity of the point changes by equal amounts in equal intervals of time. By equal changes of velocity must be understood changes of velocity that are equal in magnitude and in the same direction. When the changes of velocity in equal intervals of time are not equal, the acceleration is *variable.*

The statement that the velocity of a point is variable may refer to a change in the magnitude of the velocity, to a change in the direction of the velocity, or to a change in both. Hence we shall have three cases of acceleration to consider: (1) the acceleration of a point when the velocity of the point is constant in direction but variable in magnitude; (2) the acceleration of a point when the velocity of the point is constant in magnitude

but variable in direction; (3) the acceleration of a point when the velocity of the point is variable in both magnitude and direction.

The simplest case is when the velocity of the moving point is constant in direction and when the acceleration is constant and in the direction of the line of motion. This is illustrated by a body dropped from a height and falling in a straight line.

The *magnitude of a constant acceleration* is the magnitude of the velocity added in each unit of time, and the direction of the acceleration is the direction of the added velocity. The *unit of acceleration* is that of a point the velocity of which increases by unit velocity in unit time. When the cm. is taken as unit of length and the sec. as unit of time, the unit of acceleration is such that the velocity increases by one cm. per sec. in each second, or, briefly, one cm. per sec. per sec.

27. Motion in a Straight Line with Constant Acceleration.—In considering the motion of a point along a straight line, we take one direction along the line as positive and the opposite direction as negative and we do not need to distinguish between speed and velocity (§16). Let v_0 be the velocity of the point at the beginning of an interval of time of length t, and let v be its velocity at the end of the interval. The increase of velocity is $(v-v_0)$ and the increase per unit time is $(v-v_0)/t$. This is, therefore, the magnitude of the constant acceleration, which we shall denote by a. Hence

$$v=v_0+at \tag{1}$$

This very important equation is simply a statement that the final velocity (at the end of the time t) is equal to the initial velocity (at the beginning of t) plus the increase of velocity, and the increase of velocity is equal to the acceleration multiplied by the time.

To find how far the point travels in the time t let us consider the form of the velocity curve (§22) in the present case. The changes of velocity in equal short intervals of time are equal. Hence, in Fig. 6, the differences between each ordinate and the next in order are equal, and the velocity curve is therefore a straight line, as in Fig. 11. Draw BD parallel to OA. The whole area $OBCA$ consists of two parts, that of the rectangle

$O\ DA$ and that of the triangle BDC. OB represents the initial v ocity v_0, and we shall suppose that the figure is drawn to such a scale, that OB contains as many units of length as v_0 contains units of velocity, and that the same is true of AC, which represents the final velocity v. The height of the triangle, DC, represents in the same way the increase of velocity at. OA represents the time t, and we shall suppose that OA contains the same number of units of length as t contains units of ime. The whole area is therefore $(v_0t+\frac{1}{2}t\cdot at)$. Hence, if s is the whole distance traversed in the time t,

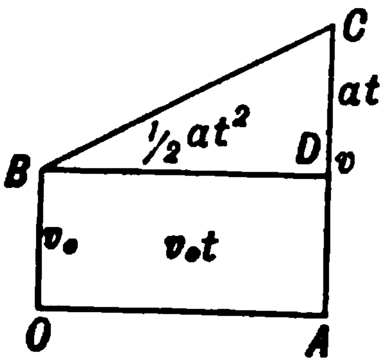


FIG. 11.

$$s=v_0t+\tfrac{1}{2}at^2 \tag{2}$$

This very important equation consists of two parts. The part, v_0t, is the distance the point would have travelled in the time t, if its velocity throughout t had remained constant and equal to the initial velocity v_0. The part $\frac{1}{2}at^2$ is the additional distance due to the acceleration, that is, the distance the point would have gone if it had started from rest with an acceleration a.

Between (1) and (2) we may eliminate t and so find an expression for the final velocity in terms of the initial velocity, the acceleration, and the distance.

$$\begin{aligned} v^2 &= v_0^2+2v_0at+a^2t^2 \\ &= v_0^2+2a(v_0t+\tfrac{1}{2}at^2) \\ &= v_0^2+2as \end{aligned} \tag{3}$$

Equations (1), (2), and (3) are of great importance.

Another expression for the area $OBCA$ is $\frac{1}{2}(AC+OB)\cdot OA$. Hence the distance is also given by the formula

$$s=\frac{v+v_0}{2}t$$

From this it follows that the average velocity, which equals the total distance divided by the time (§19), is equal to one-half of the sum of the initial velocity and the final velocity.

Equation (2) may be readily obtained by means of the Integral Calculus. The distance travelled in a short time dt when the velocity is v is vdt. Hence the whole distance, s, $=\int_0^t vdt=\int_0^t(v_0+at)dt=v_0t+\frac{1}{2}at^2$.

28. Galileo's Experiments.—The very important relations expressed by (1) and (2) were discovered by Galileo by studying the motion of falling bodies, and this discovery was the beginning of Kinetics. Before that time nothing was known as to the way in which the velocity of a body increases as it falls. Galileo thought the law of increase expressed by (1), namely, that the increase of velocity is proportional to the time, was probably correct; but the instrumental means at his command did not enable him to test it; so he deduced (2), practically by the graphical method given in §27, and then tested it. To avoid having to deal with any great velocities, such as that of a body falling vertically, he tested the rolling of a ball down an inclined plane, assuming that both motions would follow the same law. The result confirmed his formula.

29. Acceleration of Free Fall.—We shall assume as an experimental fact, discovered by Galileo, that at any one place all bodies falling freely would have the same acceleration, if it were not for the effect of air friction. The latter is very small in the case of dense solids, such as blocks of metal, falling moderate distances, and may usually be neglected. The acceleration of free fall, or the acceleration of gravity, as it is often called, is usually denoted by g. In the c. g. s. system g is about 980 cm. per sec. per sec., though slightly different at different points on the earth's surface, and in feet and seconds it is about 32.2 ft. per sec. per sec. Hence, from §27, when a body is projected vertically downward with a velocity v_0, its velocity and distance after an interval t may be found from

$$v=v_0+gt,$$
$$s=v_0t+\tfrac{1}{2}gt^2,$$
$$v^2=v_0^2+2gs$$

When the direction of projection is upward we may take upward as the positive direction, and g, being downward, will then be negative. In this case

$$v=v_0-gt, \qquad (1)$$
$$s=v_0t-\tfrac{1}{2}gt^2, \qquad (2)$$
$$v^2=v_0^2-2gs \qquad (3)$$

At the highest point $v=0$; hence from (1) we have $t=v_0/g$. Substituting this in (2), we get for the *height of ascent* $s=\frac{1}{2}v_0^2/g$.

This also follows from (3) by putting $v=0$. The time of return to the ground is got by putting $s=0$ in (2). This gives $t=2v_0/g$, showing that the whole time of rise and fall equals twice the time of ascent, or that *the time of rise equals the time of fall.* It follows from (3) that the velocity of return to the starting point, that is, when s is again zero, equals the velocity of projection in magnitude, but it is in the opposite direction. It must, however, be remembered, that these statements are true only for moderate velocities. At high velocities, such as those of a bullet, air-resistance greatly modifies the motion.

The value of g at any station of observation depends on the latitude of the station and also on the height of the station above sea level. The results of very careful experiments show that, at a station in latitude λ and at an elevation of l meters above sea level,

$$g=977.989\ (1+.0052\ \sin^2\lambda-.0000002\ l)$$

30. Motion of a Projectile.—When a body is thrown obliquely into the air, its motion may be considered as consisting of a horizontal part and a vertical part. The vertical part is subject to a constant acceleration g downward; while, since there is no horizontal acceleration (if we may neglect air-friction), the horizontal part of the motion is a constant velocity. If the magnitude of the velocity of projection is v and the direction of projection makes an angle θ with the horizontal, the velocity may be resolved into a component $v \cos \theta$ in a horizontal direction and a component $v \sin \theta$ in a direction vertically upward. If, then, x is the horizontal distance traversed in time t,

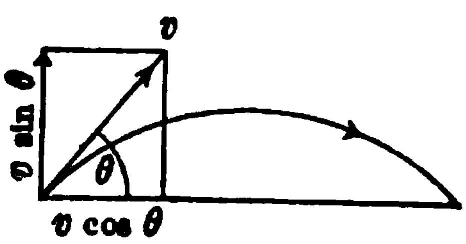


FIG. 12.—Path of a projectile.

$$x=vt \cos \theta \qquad (1)$$

and if, at the time t, the vertical distance attained is y,

$$y=vt \sin \theta-\tfrac{1}{2}gt^2 \qquad (2)$$

Thus the vertical motion is the same as that of a body thrown vertically upward with a velocity $v \sin \theta$. Hence (§29) at time $(v \sin \theta)/g$ the body will have just lost its vertical velocity and will therefore be moving wholly in a horizontal direction; and at that moment the height will be $(v^2 \sin^2 \theta)/2g$. At the time

$(2v \sin \theta)/g$ the body will have returned to its original level and the distance horizontally from its starting-point will then be $v \cos \theta \cdot (2v \sin \theta)/g$ or $(v^2 \sin 2\theta)/g$. Now, since $\sin 2\theta$ has its maximum value, unity, when 2θ is $90°$, that is, when θ is $45°$, it follows that the greatest horizontal range for a given velocity, v, of projection is v^2/g and is obtained by making the angle of projection $45°$.

If it be desired to find the constant relation that holds between x and y during the motion, the value of t taken from (1) may be substituted in (2) and we shall get

$$y = x \tan \theta - x^2 g / 2v^2 \cos^2 \theta$$

the equation of a parabola referred to axes through the point of projection. Hence the path of the projectile is a parabola.

As in the case of §29, these results are approximately correct only in the case of the moderate velocities for which air-friction is negligible. (See article on "Ballistics" Ency. Britt., 11th edition.)

31. Variable Acceleration.—When the acceleration of a point is variable, we can no longer measure it by the actual increase of velocity in any time. We may, however, divide the magnitude of the increase of velocity in any time by the time and call this the magnitude of the *average acceleration* in that time, the direction of this average acceleration being the direction of the increase of velocity. The *instantaneous value of the acceleration* is defined much as in the case of instantaneous velocity, namely, as the value to which the average acceleration approaches as the interval is diminished without limit, or

$$a = \left[\frac{\Delta v}{\Delta t}\right]_{\Delta t = 0} = \frac{dv}{dt}$$

A variable acceleration may be variable as regards magnitude or direction or both. In the following we shall consider the case of an acceleration that is constant in magnitude but variable in direction.

32. Acceleration of a Point Which Moves in a Circle with Constant Speed.—Let P be the position of the moving point at time t and P' its position at time $t + \Delta t$. At P the point is moving in the direction of the tangent PT and at P' in the direction of the tangent $P'T'$ (Fig. 13).

From any point O draw lines OQ, OQ' of equal length to represent the velocities v and v' at P and P' respectively. QQ' will

represent the velocity, Δv, added in time Δt. The triangles OQQ' and OPP' are similar. The arc PP' equals $v\Delta t$ and the chord PP' approaches equality with the arc PP' as Δt is diminished. Hence approximately

$$\frac{\Delta v}{v}=\frac{v\Delta t}{r}$$

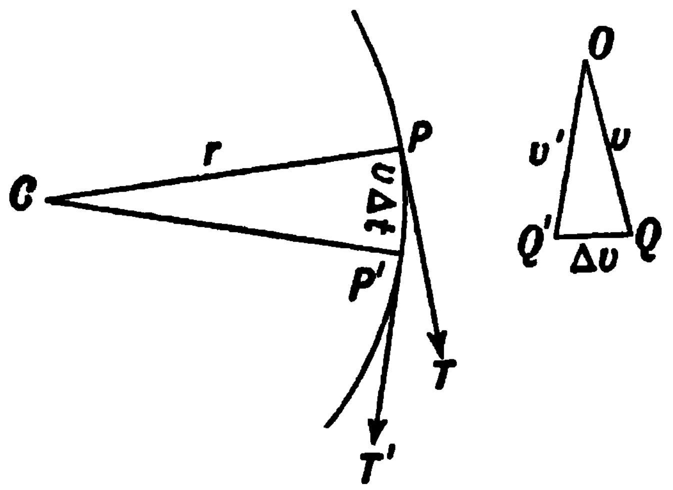


FIG. 13.

and this relation becomes more nearly exact as Δt is diminished. Hence (§31)

$$a=\frac{v^2}{r}$$

QQ' is perpendicular to PP' and in the limit it is in the direction PC. Hence *the acceleration is directed toward the center.* It is, therefore, an acceleration of constant magnitude but of variable direction.

33. Curvilinear Motion.—If in the preceding the *speed* were not constant, there would, in addition to the acceleration toward the center, be an acceleration along the tangent. The first acceleration would have the effect of changing the direction of the velocity, while the second would have the effect of changing the magnitude of the velocity, that is, the speed.

When a point moves with constant *speed* in a curve of any form, it may be regarded as moving at any moment in a circle which coincides at that point with the curve; this circle is called the circle of curvature at that point on the curve. From the radius of the circle of curvature at a point on the curve we can calculate the acceleration toward the center of the circle of curvature. If the *speed* of the point is not constant, the point must also have an acceleration along the tangent to the curve.

34. Addition of Accelerations.—A moving point may have two or more accelerations simultaneously. Thus, a man at rest on the deck of a ship which is moving with an acceleration has one acceleration, that of the ship. If he moves across the deck with an acceleration independent of the motion of the ship, he has a second acceleration. In any such case the moving body travels in some curve with a definite acceleration which is called the *resultant* of the component accelerations.

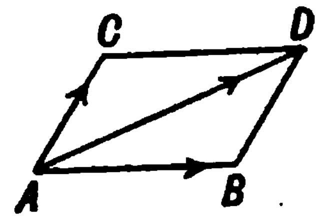


FIG. 14.

We can readily show that the resultant acceleration may be deduced from the component accelerations by the vector method

of addition, that is, by the construction of a triangle, parallelogram or polygon, the sides of which represent the separate accelerations. For let $\overline{AB}$ and $\overline{AC}$ represent two constant accelerations possessed simultaneously by a point. Since the acceleration represented by $\overline{AB}$ is constant, the change of velocity it produces in unit time is also represented by $\overline{AB}$. Similarly $\overline{AC}$ represents the change of velocity in unit time due to the second constant acceleration. The resultant change of velocity is found by completing the parallelogram $ABDC$; hence $\overline{AD}$ is the resultant change of velocity in unit time, that is, the resultant acceleration. The same method of reasoning is applicable when the accelerations are variable; the only difference being that $\overline{AB}$ and $\overline{AC}$ and $\overline{AD}$ all represent velocities that would have been added in unit time, if the accelerations remained constant that long.

35. Resolution of an Acceleration into Components.—Since two or more accelerations may be replaced by their resultant, it follows that an acceleration may be resolved into two or more components by the ordinary methods. The case in which an acceleration is resolved into two components at right angles is especially important. As an example, suppose a body rests on a smooth plane the inclination of which to the horizontal is i. If the body were not supported, it would fall with an acceleration of g. The acceleration may be resolved into a component $g \cos i$ perpendicular to the plane and a component $g \sin i$ parallel to the plane. The component perpendicular to the plane has no effect, since motion perpendicular to the plane is prevented, whereas the other component causes it to slide down the plane with an acceleration g sin i. Thus the motion down the plane may be calculated by the formulæ of §29, a being replaced by g sin i.

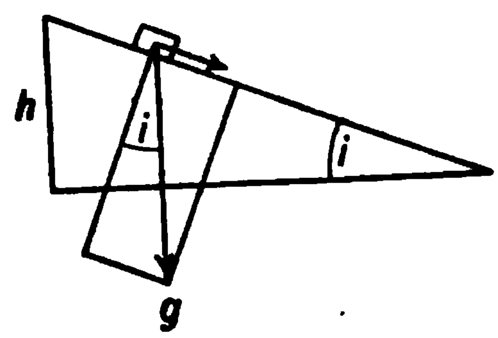


FIG. 15.—Acceleration down a plane.

From this we may deduce one result of importance. The velocity after a distance of descent s down the plane is given by

$$\begin{aligned} v^2 &= v_0^2 + 2as \\ &= v_0^2 + 2g \sin i \cdot s, \end{aligned}$$

and, if h is the distance of descent measured vertically, $h = s \cdot \sin i$. Hence

$$v^2 = v_0^2 + 2gh$$

Now this is the formula we should have been led to, if we had sought the velocity attained by a free vertical fall through a distance h. Hence the *speed attained by a body which slides without friction through a certain vertical distance is the same as if the body had fallen that distance vertically.* This does not apply to a body *rolling* down a rough plane.

DYNAMICS

Force and Mass

36. In the preceding we have considered various cases of motion without any reference to the influences that affect the motions of bodies, just as in Geometry we study lines and figures without any reference to particular bodies. We must now consider those relations between bodies on which changes of motion depend.

Isaac Newton was the first who attained clear ideas as to the relations between bodies and their motions. His treatment of the subject was founded on three fundamental principles which he called Axioms or Laws of Motion. These axioms are so simple that they are recognized as very probably true as soon as their meaning is grasped. The proof of their correctness is, however, the fact that all deductions from them are found to be verified by observation and experiment.

37. Newton's First Law of Motion.—*Every body persists in its state of rest or of uniform motion in a straight line, unless it is compelled by some force to change that state.*

This law may be divided into two parts, a *statement* and a *definition.* The statement is that any change of velocity of a body, that is any acceleration, is due to some external influence, and a body free from external influences would necessarily have a constant velocity. This was a complete denial of what had been supposed to be true up to the time of Galileo (who died in 1642, the year in which Newton was born); for, until then, it was supposed that a body free from external influences would come to rest. We cannot, of course, free any body entirely from external influences; but we can greatly diminish these influences, and with

each diminution the velocity becomes more nearly constant. The most common hindrance to steady motion is friction. A stone given a push along a rough road is quickly stopped by friction; on a smooth floor it will continue longer in motion; a well-polished stone started on smooth ice will continue in motion for a great distance. Such considerations make it seem probable that, if freed from external influences, a body would move with constant velocity; they do not, however, amount to a *proof* of the statement in the first law of motion. The proof of the law is that all of the innumerable deductions made from it and the other laws of motion are verified by experience.

The law implies a definition of **force.** This is usually given in the form "Force is whatever changes or tends to change the motion of a body" or "Force is that which produces acceleration." Thus, friction, the pull of a stretched spring, the attraction of the earth on a body, etc., are forces; when a body revolves in a circle, it has an acceleration toward the center and must, therefore, be acted on by some force. What exerts a force on a body is, of course, some other body. Thus the friction opposing the motion of a vehicle is due to the earth and the pull of a spring is due to the spring, etc. The word "force" is therefore a name which we give to that influence of one body on another by which the first changes the motion of the second.

The property a body has of tending to persist in its state of motion or of rest is called Inertia.

38. The Mass of a Body.—Common experience shows that, when a given force is applied to a body, the magnitude of the acceleration depends on some property of the body. Thus, a horizontal spring kept stretched to a definite length, say one foot, will apply a definite force to the body to which it is attached. If attached to a cubic foot of lead supported without friction, it will produce a certain acceleration; but the acceleration will be different, if a cubic foot of wood be substituted for the lead. The difference is not due to the difference in the weights of the bodies, since weight is a force that acts vertically and does not affect the horizontal motion of the bodies. The difference is due to what we call the masses of the bodies.

To attach a definite meaning to the word **mass** we must define what is meant by the ratio of the masses of two bodies. *The ratio*

of the masses of two bodies is the inverse of the ratio of the accelerations that a given force imparts to the bodies when applied to them in succession. For example, if a body A acted on by a certain force receives twice the acceleration that a second body B receives when acted on by the same force, the mass of A is half as great as the mass of B and so for other ratios. Hence, if we adopt, as we presently shall, a certain body as a body of unit mass, the mass of any other body becomes definite.

In the above we have defined the ratio of two masses by means of the ratio of the accelerations imparted to them by *some particular force.* This at once suggests the question: Would the ratio be found different if some other force were used in the test? If so, the word mass as applied to a body would have no definite meaning except in relation to a particular force. But, as a matter of fact, *the ratio is found to be the same, no matter what may be the force chosen for the test.* This is a very important statement, but we do not need to state it as a separate fundamental principle since it is included in Newton's Second Law of Motion as will readily be seen later from the formula for that law (§42). The fact that *bodies have definite masses,* the same no matter what their accelerations or the forces acting on them, was one of Newton's most important discoveries.

We shall see later (§570) that there is good reason to believe that at immensely great velocities the mass of a body may depend appreciably on its velocity.

Some persons find difficulty in accepting the above definition of the ratio of two masses, because they cannot see an easy means of applying it directly to comparing masses. It is, however, not given as a practical method of comparing masses; but it leads, as we shall see later, to a very practical method (§42).

39. Units of Mass.—The unit of mass chiefly employed in Physics is the *gram,* which is defined as one one-thousandth of the mass of a block of platinum kept at Sèvres, near Paris, and known as the *Kilogram prototype.* Fractions and multiples of the gram in frequent use are named as follows:

Milligram	=.001 g.	Kilogram	=1000 g.
Centigram	=.01 g.	Metric ton	=1,000,000 g.
Decigram	=.1 g.		=1000 kg.

In English-speaking countries the *pound* is, for commercial and industrial purposes, used as unit of mass. It is defined as the mass of a certain block of platinum kept at the Exchequer in

London. It is worth remembering that 1 kgm. = 2.20 lbs. approximately and that 1 pound = 454 gms. approximately.

40. Ratio of Forces.—Different forces applied to a body give it different accelerations. For example, if a heavy body be hung from the ceiling by a cord and a horizontal cord be attached to it, a small pull will start it slowly, while a stronger pull will start it more rapidly. Or, if a horizontal spiral spring, kept stretched to a definite length, were applied to a body supported with very little friction on a horizontal table, a definite acceleration would be produced. If this experiment were repeated with the spring stretched to a different length, a different acceleration would result. These illustrations would be somewhat difficult to carry out accurately, but they will help to make clear the following definition of the ratio of two forces, and from this we shall be able to deduce a more accurate method of finding the ratio, either by calculation (§42) or by static experiments (§52).

The ratio of two forces is the ratio of the accelerations they can impart to a given body. For definiteness, we shall suppose that the body referred to is one of unit mass. If now we take any force as unit force, the magnitude of any other force becomes definite. For simplicity, we shall usually take *as unit force that force which, acting on unit mass, gives it unit acceleration.* A force which gives unit mass two units of acceleration will then be a force of two units, and so on.

In the above we have defined the ratio of two forces by the ratio of the accelerations they impart to some particular body. This definition would not be of much value, if the ratio obtained depended on the particular body chosen. As a matter of fact, the ratio of two forces, as defined above, is the same, no matter what body is chosen for the test. This statement, while very important, does not need to be stated as a separate fundamental principle, since it is included in Newton's Second Law of Motion, as will readily be seen later from the formula for that law (§42).

41. Momentum.—Every one is aware that certain properties of moving bodies depend on mass and velocity conjointly. Thus, the length of time required by a locomotive to start a train depends on both the mass of the train and the velocity to be imparted to it, and the same is true of stopping it. Hence we find it convenient to define a property depending on mass and velocity conjointly. *Momentum* is defined as *the product of mass*

and velocity. Since the velocity has direction as well as magnitude, while the mass has magnitude only, the momentum of a body is a vector quantity, the direction of which is that of the velocity. (What we now call momentum Newton called *quantity of motion* as distinguished from *rate of motion* or velocity.)

When the velocity of a body changes, its momentum also changes. Since the mass of the body is constant, any change in the momentum of a body must be due to a change of its velocity, and the change of momentum must equal the product of the mass and the change of velocity. Hence, when the momentum of a body is changing, the rate of change of momentum equals the product of the mass and the rate of change of velocity, that is, the product, ma, of the mass m and its acceleration a.

42. Newton's Second Law of Motion.—*The rate of change of the momentum of a body is proportional to the force acting on the body and is in the direction of the force.*

To reduce this statement to a mathematical formula, let us suppose that a force F_1 acting on a mass m_1 gives it an acceleration a_1, and that a force F_2 acting on a mass m_2 gives it an acceleration a_2, and so on for any number of forces and masses. Then

$$F_1 : F_2 \ldots :: m_1a_1 : m_2a_2 \ldots$$

If k be used to denote the constant ratio of F to ma

$$F = kma$$

This is the general formula for Newton's Second Law. The constant k is a number the magnitude of which depends on the unit chosen for F, since we have already chosen certain units for m and a.

Since, as Galileo found (and as Newton and Bessel proved more completely), all bodies fall with the same acceleration (allowance being made for air friction), it follows from the above formula that *the masses of bodies are proportional to their weights.* This is the principle of the common balance, by which the masses of bodies are compared by comparing their weights.

43. Units of Force.—(1) *Absolute Units.*—Calculations by means of the formula for Newton's Second Law of Motion are much simplified when the unit of force is so chosen that k is 1. Such a unit of force is called an absolute unit of force.

Since F is to be 1 when m and a are each 1, an absolute unit of force is that force which, acting on a body of unit mass gives it unit acceleration. Hence

$$F = ma \text{ in absolute units.}$$

If the gram be taken as unit of mass and the cm. per sec. per sec. as unit of acceleration, the absolute unit of force is *that force which, acting on a body of one gram mass, gives it an acceleration of one cm. per sec. per sec.* and is called a *dyne.*

If the pound be taken as unit of mass and the foot per sec. per sec. as unit of acceleration, the absolute unit of force is that force which, acting on a body of one pound mass, gives it an acceleration of one foot per sec. per sec. and is called a *poundal.*

(2) *Gravitational Units.*—The weight of a body of unit mass is, for many purposes, a convenient unit of force; but, when it is chosen, the value of k is not 1. For allow a body of unit mass to fall: the acceleration $a = g$, while $F = 1$ and $m = 1$. Hence, substituting in $F = kma$, we get $1 = kg$, and therefore $k = 1/g$. Hence

$$F = \frac{m}{g}a \text{ in gravitational units.}$$

The only gravitational unit we need consider is the weight of a pound. With the foot per sec. per sec. as unit of acceleration, the value of g is 32.2 approximately but varies with the locality. Formulæ derived from the above will always have m/g where m only would appear in absolute units.

Engineers prefer to write W, the number of pounds weight in the weight of a body, instead of m, the number of pounds mass in the mass of a body. The two are equal numerically.

44. Newton's Second Law (Continued).—The statement of the second law of motion is so brief that some things implied in it might easily escape notice:

1. In the statement of the law the rate of change of momentum of a body is spoken of, without any reference to whether the body starts from rest or is initially in motion. Hence it is implied that *the effect of a force applied to a body is independent of the state of motion of the body when the force begins to act.* For example, gravity is a force that acts vertically downward. When a body is dropped from a height, the force of gravity gives it a certain

acceleration downward; if the same body be started downward with a certain velocity, its acceleration downward will be the same as when the body is simply dropped, and the same will be true if the body be given an initial velocity upward or in any direction. It is found possible to play games of ball or cricket on a moving steamship; the effect of throwing the ball with a certain force or striking it with a bat is the same as when the steamship is at rest.

2. The law states how a force will affect the motion of a body, but it makes no reference to whether some other force is acting on the body at the same time or not. Hence it is implied that *each force produces its own effect independently of the simultaneous action of any other force;* and, when several forces act on a body, we may calculate the acceleration produced by each as if the other forces did not exist, and then add the accelerations to find the whole effect of all the forces. This very important principle is sometimes called that of the *independence of forces.*

45. Impulse of a Force.—The product of a force and the time during which it acts is called the *impulse* of the force. When a force F acts on a mass m for time t, from the formula for the Second Law of Motion, by multiplying both sides by t, we get:

$$Ft = mat$$

Now at is the increase of velocity produced, and this, multiplied by m, is the increase of momentum. Hence *the impulse of a force equals the momentum produced by it.* If the body, starting with a velocity v_0, has at time t a velocity v,

$$Ft = mv - mv_0$$

46. Newton's Third Law of Motion.—*Action and reaction are equal and opposite.* In the statements of the first and second laws of motion forces acting on bodies are spoken of, but nothing is said as to what exerts force. This lack is supplied by the third law.

The action and reaction here referred to mean *force and counter-force.* The meaning of the statement is that force on any one body is exerted by some other body, and this other body itself experiences an equal and opposite force exerted by the first body, the line of action of both forces being the line joining the two bodies.

In many cases the truth of this law will be recognised as being evident. For example, when one presses his two hands against each other, it will be admitted that the hands, if at rest, press equally in opposite directions. If one hand be pressed against a wall, the same must still hold, since the wall merely takes the place of the other hand in the first illustration. But the case is not so clear when a hand is pressed against an obstacle that moves. How, it is sometimes asked, can there be motion produced if the forces are equal and opposite? The answer is that *the two forces spoken of do not act on one body;* there is one force exerted by the hand on the obstacle, and the obstacle yields unless restrained by some other force; the reaction is the back pressure of the body *on the hand*, not a force acting on the body.

Consider, also, the forces that come into play when a horse of mass m_1 pulling on a horizontal rope of mass m_2 draws a block of mass m_3. Here there are four pairs of actions and reactions. In the first place, the horse pushes against the ground and the reaction of the ground is an equal and opposite push. Let the magnitude of this horizontal action and reaction be F_1. Secondly the horse exerts a forward pull, of magnitude say F_2, on the rope and the reaction of the rope is equal and opposite. The rope exerts a horizontal force on the block and the block exerts an equal and opposite reaction, the magnitude of each being F_3. Finally, there is the action and reaction between the block and the ground; let the horizontal component of this have a magnitude F_4. If there is an acceleration a, as there must be to begin the motion, F_1 is greater than F_2 by m_1a, F_2 is greater than F_3 by m_2a, and F_3 is greater than F_4 by m_3a. Thus F_1 exceeds F_4 by $(m_1+m_2+m_3)a$, and this is, therefore, the total backward push on the ground. When the motion has become constant, $a=0$ and all the forces mentioned are of equal magnitude.

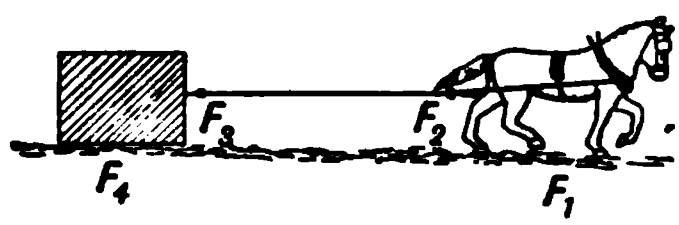


FIG. 16.—Four pairs of actions and reactions.

Since a force is always accompanied by a counterforce, the two are parts or different aspects of one inseparable whole, and the two together constitute what is called a *stress*. Thus *every force is the partial aspect of some stress*, just as a purchase and a sale are partial aspects of an exchange.

47. Force Required for Motion in a Circle.—When a particle revolves in a circle, it has an acceleration toward the center equal to v^2/r (§ 32), where v is the magnitude of the velocity (*i.e.*, the speed) and r is the radius. To cause this acceleration there

must be a force directed toward the center, and, according to Newton's Second Law, this *centripetal* force F must be such that

$$F = m\frac{v^2}{r}.$$

Against this force the particle will exert an equal and opposite reaction on the body that exerts the force toward the center. If, for example, the particle be attached by a string to the finger, the reaction will be a force acting on the finger and will be in a direction outward along the radius. This reaction is called a *centrifugal force*. Thus the centrifugal force is *not a force acting on the moving particle, but a reaction, exerted by the particle, on the other body that exerts the force toward the center.* (That the above formula also applies to the motion of a *body* is shown in §101.)

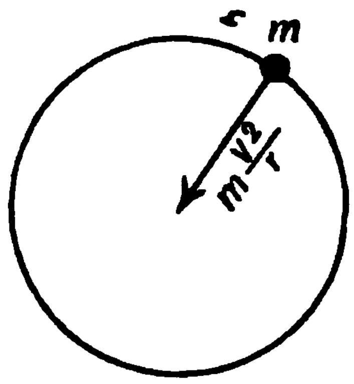


FIG. 17.—A particle moving in a circle is acted on by a force toward the center.

Illustrations of the above are very numerous and a few may be mentioned. *Drops of water* are thrown off tangentially from a rapidly moving bicycle or carriage wheel, owing to the fact that there is not a sufficient force toward the center acting on them, and they therefore move off on a tangent, in accordance with the first law of motion. A *train rounding a curve* presses outward on the rails, and the resultant of this force and the vertical weight of the train is a force inclined to the vertical. Since it is desirable that the whole force should be perpendicular to the sleepers, the outer rail is raised. In the *Centrifugal drier*, used in laundries and sugar refineries, the material to be dried is placed in a perforated cylinder rotating about its axis which is vertical; the drops of water, not being held by a force directed to the center, escape through the perforations. In the *Centrifugal cream-separator*, which is a rotating vertical cylinder, both the milk and the cream tend to move as far from the axis as possible; but the milk, being the denser, exerts the more powerful tendency and therefore occupies the parts of the vessel farthest from the axis. The *flattening of the earth* at its poles is due to its axial rotation; if at rest it would be spherical; but, being in rotation, it bulges at the equator to such an extent that the restoring forces due to gravitational attractions supply the requisite force toward the center. The higher the speed of *belting* the less it presses on a pulley and the more liable, therefore, it is to slip; for more of the tension of the belting is called on to supply the requisite force toward the center. Watt's *governor for a steam-engine* consists of a pair of balls whirled around a vertical spindle at a rate proportional to the speed of the engine; when this speed exceeds the desired limit the outward movement of the balls acts on a steam-valve so as to decrease the speed of the engine.

Resultant of Forces.—Equilibrium

48. Composition of Forces.—Two or more forces may act on a body at the same time. For example, a body falling because of the attraction of the earth may be drawn horizontally by a stretched spring or blown by wind pressure. In such cases each force produces an acceleration independently of the action of the other forces (§44), and the body travels in some path with a definite acceleration, which is the resultant of the accelerations produced by the separate forces.

The resultant of two or more forces is defined as the single force which will produce the resultant acceleration. The resultant of any number of forces which act on a particle can be found by vector addition, that is by a triangle, parallelogram, or polygon construction. For a force has a certain magnitude and a certain direction and is, therefore, a vector quantity. Hence any number of forces acting on a particle may be represented by lines drawn from a point. Let $\overline{AB}$ and $\overline{AC}$ represent two forces F_1 and F_2, acting on a particle. Complete the parallelogram $ABDC$. By the Second Law of Motion the accelerations produced by F_1 and F_2 are in the directions of and proportional to $\overline{AB}$ and $\overline{AC}$, and the resultant acceleration must, therefore, be represented by $\overline{AD}$; and, since the resultant force is the force that will produce the resultant acceleration, it must be in the direction $\overline{AD}$. If, now, we denote the accelerations produced by F_1 and F_2 by a_1 and a_2 respectively and if the resultant force and acceleration be denoted by F and a respectively, by the Second Law of Motion

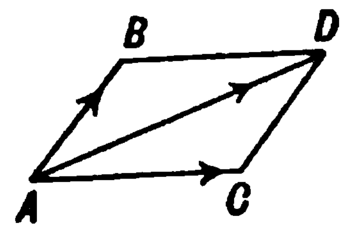


FIG. 18.

$$\begin{aligned} F : F_1 : F_2 &:: ma : ma_1 : ma_2 \\ &:: a : a_1 : a_2 \\ &:: AD : AB : AC \end{aligned}$$

Hence $\overline{AD}$ represents the resultant force on the scale on which $\overline{AB}$ and $\overline{AC}$ represent the separate forces. This very important result, called the **Parallelogram of Forces**, is usually stated as follows:

If two forces acting on a particle be represented by two lines drawn from a point and if a parallelogram be drawn with these two

lines as sides, the resultant will be represented by the diagonal that passes through the point.

Since, then, we may add two forces by the parallelogram method or by the triangle method (which is essentially the same), we may in the same way add a third to the resultant of these two and so on. Hence the polygon method of addition applies to forces acting on a particle.

Let θ be the angle between the directions of the forces F_1 and F_2. Then, as in the case of velocities and accelerations,

$$F^2 = F_1^2 + F_2^2 + 2F_1F_2 \cos \theta$$

49. Resolution of a Force into Components. Since two or more forces acting on a particle can be replaced by a single force called their resultant, a single force can be replaced by any two or more forces which, added geometrically, give the single force. This is called the *resolution* of a *force* into *components*.

The most important case practically is when the components are at right angles to each other. When a single force is resolved into two components (Fig. 19), the components and the force resolved must be in the same plane. When the two components are

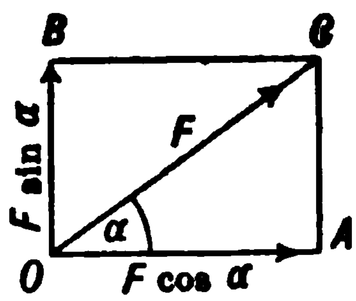


FIG. 19.

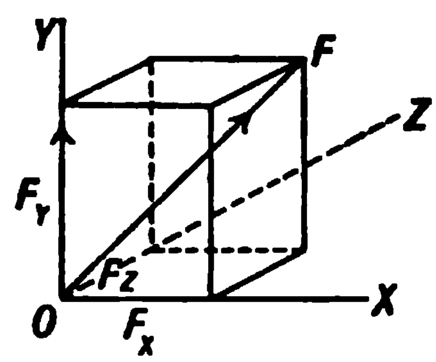


FIG. 20.

at right angles, the component that makes an angle α with the whole original force F has a magnitude $F \cos \alpha$, and the other component is $F \sin \alpha$. The agreement as regards the signs of angles noted in §25 applies to the present case.

A force F may also be resolved into three components in three directions at right angles to each other. All that is necessary is to construct a right-angled parallelopiped with the line representing F as diagonal (Fig. 20) and with edges in the three rectangular directions. If the three directions be taken as axes of x, y and z and if the components be denoted by F_x, F_y, F_z respectively, we shall have

$$F^2 = F_x^2 + F_y^2 + F_z^2$$

50. Illustrations of the Resolution of a Force into Components.—1. The force of gravity on a body of mass m acts vertically downward and, in absolute units, equals mg. If a body is not free to move vertically, but is free to move in some other direction, the only part of gravity that can affect the motion is the component in that direction. For instance, if a body (Fig. 21) be on a smooth plane inclined at an angle i to the horizontal, the force of gravity, mg, may be resolved into a component $mg \sin i$ down the plane and a component $mg \cos i$ perpendicular to the plane. The latter component will produce pressure on the plane but will not affect the motion down the plane, which will depend only on the former component, $mg \sin i$. If the plane be not perfectly smooth, there will also be a force of friction, say F, parallel to the plane, and the resultant force down the plane will be $(mg \sin i - F)$.

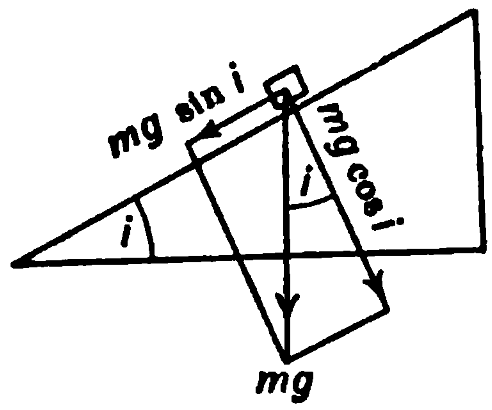


FIG. 21.

2. A sail-boat (Fig. 22) effects a double resolution of the wind pressure. The component of the wind pressure W parallel to the plane of the sails has very little effect; the component, say F, perpendicular to the sail is the effective component. Again, F may be resolved into a component perpendicular to the keel and a component f parallel to the keel. The former produces a small sidewise motion or lee-way, while the latter, being in the direction in which the boat is most free to move, is the effective component.

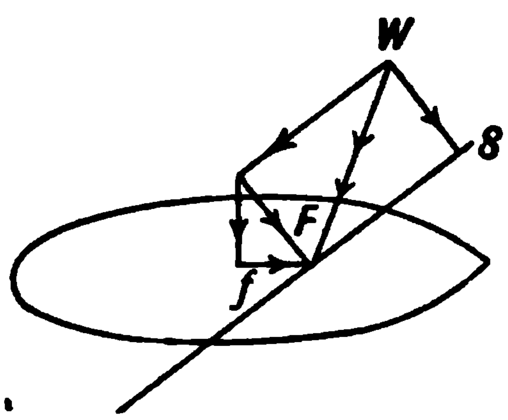


FIG. 22.

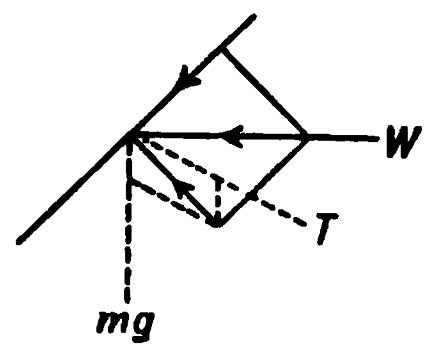


FIG. 23.

3. In the case of a kite (Fig. 23) the component of the wind pressure parallel to the surface of the kite has no effect; the component perpendicular to the kite, when the kite has risen to the proper level, is equal and opposite to the resultant of the pull, T, of the cord and the weight, mg, of the kite.

51. Analytical Method of Compounding Forces.—A simple and general method of finding the resultant of a number of forces in a plane is to resolve each in two directions at right angles and then add all these components. Thus, let F_1, $F_2 \cdots$ be the forces acting on a particle at O. Take any two convenient rectangular directions, Ox and Oy, and let the angles F_1, $F_2 \cdots$ make with Ox be α_1, $\alpha_2 \cdots$ respectively. Then F_1 is equivalent to $F_1 \cos \alpha_1$

along Ox and $F_1 \sin \alpha_1$ along Oy and so for the other forces. Let the sum of the components along Ox be denoted by X and the sum of the components along Oy by Y. Then

$$X = F_1 \cos \alpha_1 + F_2 \cos \alpha_2 + \cdots = \Sigma F \cos \alpha$$
$$Y = F_1 \sin \alpha_1 + F_2 \sin \alpha_2 + \cdots = \Sigma F \sin \alpha$$

We have thus replaced the forces $F_1, F_2 \cdots$ by X along Ox and Y along Oy. The resultant of X and Y is the resultant of F_1, F_2, etc. Let the magnitude of the resultant be R and let it make an angle θ with OX. Then

$$R^2 = X^2 + Y^2$$
$$\tan \theta = \frac{Y}{X}$$

These formulæ give the magnitude and the direction of the resultant. In using this method it must be remembered that, when we substitute for each angle α its numerical value, we must call the angle positive if it is measured in the direction regarded as positive, say the counter-clockwise direction; if measured in the opposite direction it must be regarded as negative.

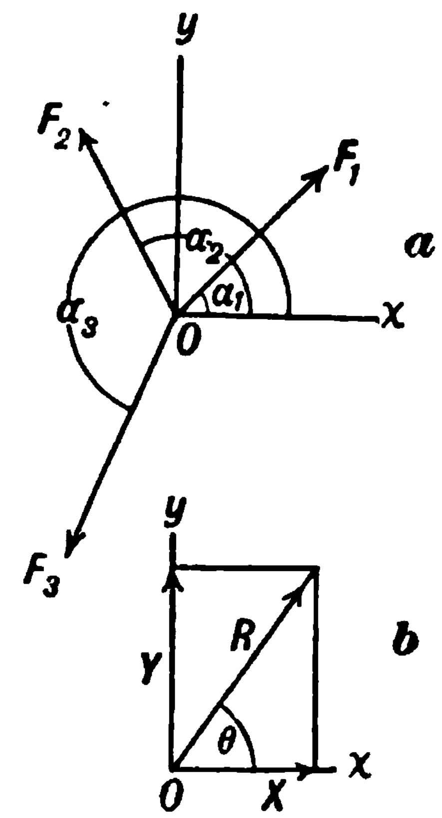


FIG. 24.—Analytical method of compounding forces.

The angle θ that the resultant makes with Ox is found from its tangent. When the tangent is positive, it shows that the angle is between 0 and 90° or between 180° and 270°. To decide between these two, note that the signs of the values of X and Y must be either both positive or both negative, since the tangent is positive. If both are positive, θ is between 0° and 90°; if both are negative, it lies between 180° and 270°. The reader should have no difficulty in completing the reasoning for the case in which $\tan \theta$ is negative.

When X and Y are both zero, that is, when the sum of the components in each of two directions at right angles is zero, R is also zero. Conversely, when R is zero, X and Y must also each be zero, since the square of a number cannot be negative.

When the forces to be compounded are not all in one plane, we may take three directions, Ox, Oy, Oz, at right angles and resolve each force into components in these three directions. Denote the sum of the components along Ox by X, that along Oy by Y, and that along Oz by Z and let the resultant be R. Then

$$R^2=X^2+Y^2+Z^2$$

If $X=0$, $Y=0$ and $Z=0$, then $R=0$. The converse is also true, since X^2, Y^2 and Z^2 must be either positive or zero.

52. Equilibrium of Forces Acting on a Particle.—*When the resultant of the forces acting on a particle is zero, the forces are said to be in equilibrium,* that is, in a state of balance, so that they do not change the motion of the particle.

When two equal and opposite forces act on a particle, they are in equilibrium, for their resultant is zero. Conversely, if two forces are in equilibrium, they must be equal and opposite, for otherwise their resultant could not be zero.

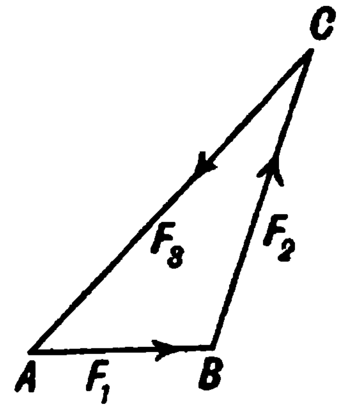


FIG. 25.

When *three forces acting on a particle are in the direction of and proportional to the sides of a triangle taken in order, they are in equilibrium.* For if the three forces F_1, F_2, F_3 be in the direction of and proportional to $\overline{AB}$, $\overline{BC}$, $\overline{CA}$, the resultant of F_1 and F_2 will be represented by $\overline{AC}$, and the resultant of forces represented by $\overline{AC}$ and $\overline{CA}$ is zero. The converse of this proposition

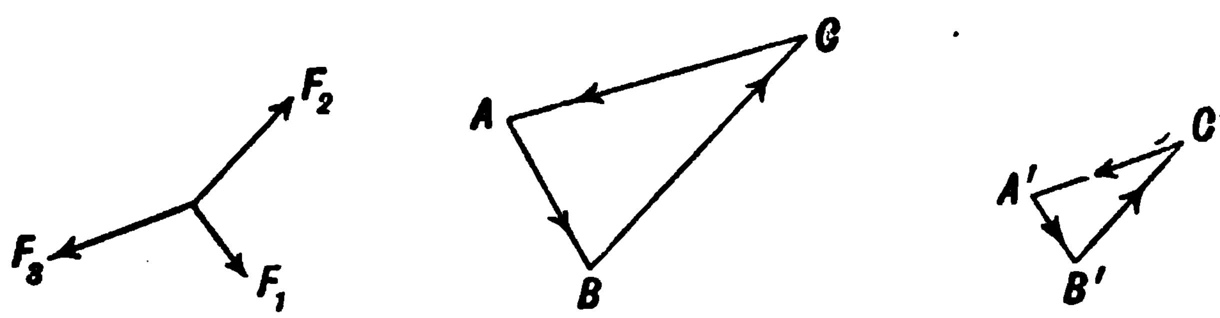


FIG. 26.

is that, *if three forces acting on a particle be in equilibrium, and if any triangle be drawn with its sides respectively in the directions of the forces, the forces will be proportional to the sides of this triangle.* To prove this let us suppose that $\overline{AB}$ and $\overline{BC}$ (Fig. 26), are any two lines in the direction of and proportional to two of

the forces F_1 and F_2. Then a force represented by $\overline{AC}$ is equivalent to F_1 and F_2 taken together. Hence, since the forces are in equilibrium, the third force F_3 must be in the direction of and proportional to $\overline{CA}$. Now *any other triangle* such as $A'B'C'$, with sides in the directions of F_1, F_2, F_3 respectively, that is, in the directions of $\overline{AB}$, $\overline{BC}$, $\overline{CA}$, respectively, must be similar to ABC. Hence its sides must be proportional to the sides of ABC, that is, to F_1, F_2, F_3 respectively. This converse proposition is very important, for, when we know the directions of three forces that are in equilibrium, we can find the relative magnitudes of the forces by constructing a triangle with its sides in the directions of the forces.

When *any number of forces* acting on a particle are in the directions of and proportional to the sides of a closed polygon taken in order, they are in equilibrium; for the resultant is zero. The converse of this proposition, for more than three forces, is not true; for polygons are not necessarily similar when their respective sides are parallel.

When *any number of forces* are such that the sum of their components in each of three directions at right angles is zero, they are in equilibrium. This is evident from §51, for, when X, Y and Z are all zero, R must also be zero. Conversely, when any number of forces are in equilibrium, the sum of their components in *any* direction equals zero; for we may take this direction as one of three at right angles; and, since R is zero, the sum of the components in each of these directions is zero (§51).

Work and Energy

53. Work.—When a force acts on a body, *the product of the force by the distance through which it acts in the direction of the force* is, as we shall see later, a very important quantity and is called the *work* performed by the force. Thus, when a force applied to a heavy body raises it a certain vertical distance, work is performed by the force, the amount of the work being the product of the force and the distance of ascent; and, when a horizontal force draws a body horizontally, the work is the product of the force and the horizontal distance.

The phrase "in the direction of the force" that occurs in the

definition of work should be carefully noted. When there is no motion in the direction of a force, no work is performed by that force. For instance, a travelling crane may by its chains exert an enormous force in sustaining a heavy body and it may move the body through a great distance horizontally, but the force exerted by the chains will do no work if there is no vertical motion. If a force F acts constantly on a body while the body moves a distance AB which is not in the direction of the force, to get the work performed we must take the projection of AB on the line of action of the force and multiply the projection by the force. If θ is the angle between AB and the direction of the force, the projection, AC, of AB on the line of the action of the force is $AB \cos\theta$ and the work performed is $F \cdot AB \cos\theta$. This at once suggests another method of calculating the work performed, for $F \cdot AB \cos\theta$ is the same as $F \cos\theta \cdot AB$ and $F \cos\theta$ is the

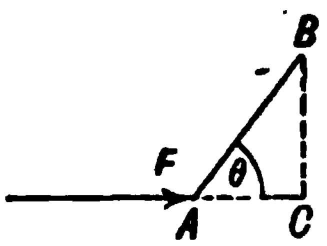


Fig. 27.

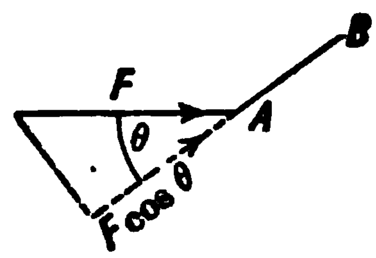


Fig. 28.

component of F in the direction of AB. Thus the work performed by a force is also the product of the total distance by the component of the force in the direction of motion.

Work, or the product of force and distance, must be carefully distinguished from the impulse of a force, which is the product of a force and the time during which it acts (§45). Given a force and the distance through which it acts, we do not need to know the time in order to calculate the work.

54. Positive and Negative Work.—Forces always exist in pairs of equal and opposite forces (§46). Hence, when a force applied to a body does work by moving the body in the direction of the force, it must at the same time overcome an opposing force or reaction. The applied force in this case does positive work, since the motion is in the direction of the applied force. This work is done against the reaction, or we may say that the reaction does negative work, since the motion is in the opposite direction to the reaction.

The nature of the reaction is different in different cases. A

horse attached to a wagon is doing work against the force of friction when the wagon is moving uniformly, and the force of friction does negative work. In starting the wagon into motion the horse does work against the inertia of the wagon and also of the horse, in addition to the work it does against friction. When a body is moving in one direction and a force is suddenly applied to it in the opposite direction, the body does positive work against the force, which in this case does negative work.

55. Units of Work.—The unit of work is *the work done by the unit force in acting through unit distance.* When the dyne is taken as unit of force and the cm. as unit of length, the unit of work is that performed by a dyne acting through a cm. and is called an erg. Since this is a very small unit, a multiple of it, namely 10,000,000, (or 10^7) ergs, is frequently used and is called a joule.

When the weight of a pound is taken as unit of force and the foot as unit of length, the unit of work is the work done by a force equal to the weight of one pound when it acts through one foot and is called a *foot-pound.* The work done by a poundal acting through a foot is called a foot poundal.

56. Diagram of Work.—When a force is constant, to find the work it does we multiply the magnitude of the force by that of the displacement; but, when a force is variable, some other method has to be adopted. One way is to divide the whole displacement up into small parts and multiply each small part by the force at the middle of the small displacement and then add all the products. Stated briefly, $W = \Sigma F . \Delta s$. By taking the parts small enough we may get the work as accurately as may be desired. A graphical method is often preferable. It is entirely similar to the method used in finding the distance a point travels when it has a variable velocity (§22). Let OA be a line that represents, on some scale, the whole displacement measured in the direction of the force. Divide OA into a very large number of very small equal parts. At O erect a perpendicular OB to represent, on some scale, the force at the beginning of the first part; erect similar perpendiculars to represent the magnitude of the force

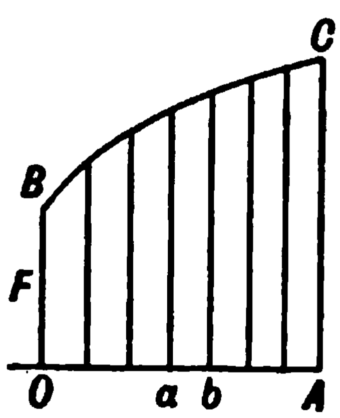


FIG. 29.—Diagram of work.

at the beginning of the other parts, and through the ends of these perpendiculars draw a smooth curve *BC*. If we calculated the work done in a small displacement *ab* by taking for the force its value at the beginning of *ab*, the result would be too small; and, if we made the calculation by taking the value of the force at the end of *ab*, the result would be too large, and similarly for all the other intervals. By continuing the reasoning as in §22, we find that the actual work done is represented by the area *OBCA*.

Thus, to find the whole work, we need only to measure the area of the figure and then allow for the scale on which it is drawn. If each unit of length along *OA* stands for *m* units of length in the displacement, and if each unit of length along *OB* stands for *n* units of force, each unit of area will stand for *mn* units of work, and the whole area multiplied by *mn* will give the whole work.

When the curve of force is a straight line the area may be readily calculated. For example, let us calculate the work done in stretching a spring. In this case it is known that the force that is needed to keep a spring stretched is proportional to the amount of the stretch or increase of length (provided this be not so great as to permanently lengthen the spring). Hence, if the spring is stretched by an amount x, the force applied to it is kx, where k is a constant and is evidently equal to the force required to produce unit increase of length. If, then, a curve be drawn with the values of kx as ordinates and the values of x as abscissæ, this curve will be a straight line (Fig. 30), which will pass through the origin, since kx is zero when x is zero. To find the work done in increasing the amount of the stretch from x_1 to x_2 where $OL=x_1$ and $ON=x_2$ we must find the area $PLNQ$. Now this is equal to $\frac{1}{2}LN(PL+QN)$. Hence the work done is $\frac{1}{2}(x_2-x_1)\ (kx_2+kx_1)$ or $(\frac{1}{2}kx_2^2-\frac{1}{2}kx_1^2)$. This is also the work the spring will do in contracting, since at each step the force of contraction is equal to the force required to stretch. If the initial stretch be zero, $x_1=0$, and the work required to stretch by the amount x_2 is $\frac{1}{2}kx_2^2$. While we have referred especially to the force exerted by a spiral spring, the above proof and formula evidently apply to the work done by any force that is

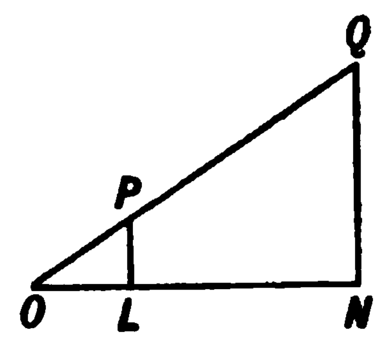


FIG. 30.

proportional to displacement. These we shall find later are very numerous.

57. Power or Activity.—The *rate* at which an agent works, or the number of units of work performed per unit time, is called the *power* or activity of the agent. In c. g. s. units the unit of activity is that of an agent that does one erg per second. As this unit is extremely small, the unit employed for most scientific purposes is 10^7 ergs per second, or one joule per second, and is called the *watt;* a still larger unit is the *kilowatt* which equals one thousand watts.

The unit largely employed for engineering purposes is the *horse-power,* which is the power of an agent that does 550 foot pounds per second or 33,000 foot pounds per minute. One horse-power equals 745.8 watts.

58. Kinetic Energy.—It is often necessary to find the relation between the work done by a force and the velocity of the body to which it is applied, as, for example, in considering the motion of a train. As the simplest case suppose a body of mass, moving with a velocity v_0 at the beginning of an interval of time t, to be acted on by a single force F in the direction of the velocity, and let the velocity at the end of the time be v and the distance traversed be s. Then from §§27 and 45

$$s=\tfrac{1}{2}(v+v_0)t \qquad (1)$$

and

$$Ft=m(v-v_0) \qquad (2)$$

Hence

$$Fs=\tfrac{1}{2}mv^2-\tfrac{1}{2}mv_0^2 \qquad (3)$$

One-half the product of the mass of a body and the square of its velocity is called the *kinetic energy* of the body, or, briefly,

$$K.\ E.=\tfrac{1}{2}mV^2$$

We may, therefore, state the above conclusion thus:

Work done on body = gain of kinetic energy;

but it must be remembered this is only for the case in which the force acts on a body which is otherwise free. It is, however, true in all cases, if F stands for the *resultant* of all the forces acting on the body.

We may also reverse the circumstances and enquire what work a body in motion can do, if it meets an opposing force and is

brought to rest. Suppose that it exerts a constant force F and does work Fs. Then the opposing force, that is the force applied to the body, will be $-F$. Making this change in (2) we get

$$Fs = \frac{1}{2}mv_0^2 - \frac{1}{2}mv^2$$

Thus the work done *by the body* against the resistance is equal to the *loss* of kinetic energy of the body.

If the motion continue until the body is brought to rest, v will then be zero, and we shall have the result that *the initial kinetic energy of the body is the work it can do before it is brought to rest.*

It should be noticed that, in the above, v and v_0 stand for the *magnitudes* of the respective velocities, *i.e.*, the speeds (§16). The kinetic energy of a body depends on the square of the magnitude of its velocity and is the same no matter what the direction of motion, that is, kinetic energy is a scalar quantity; to the kinetic energy of one body we may add the kinetic energy of another body and the sum will be the total kinetic energy of both bodies.

Since a force does no work when it is always at right angles to the direction of motion, it follows that, when a body is acted on by a single force at right angles to the direction of motion, the kinetic energy of the body remains constant. Thus, when a body rotates in a circle under the action of a single force directed toward the center, the force does no work and the kinetic energy of the body is constant.

Kinetic energy and work are equivalent quantities; hence the units of kinetic energy are the same as the units of work.

59. Kinetic Energy and Gravity.—The force of gravity on a body is, for small distances above the surface of the earth, a constant force. If a body at a height H above the earth's surface has a velocity v vertically downward, when it has fallen so that its distance above the surface is h, gravity will have done an amount of work $mg(H-h)$; and, if the velocity of the body be then V, its kinetic energy will have increased from $\frac{1}{2}mv^2$ to $\frac{1}{2}mV^2$. Hence

$$mg(H-h) = \frac{1}{2}mV^2 - \frac{1}{2}mv^2$$

If on the other hand the body be projected upward with a velocity V from a height h, it will be opposed by the force mg, and the work it will do against gravity, in rising to a height H, will be

$mg(H-h)$. If its velocity at the height H be v, its loss of kinetic energy will be $(\frac{1}{2}mV^2-\frac{1}{2}mv^2)$. Equating the work done against gravity to the loss of kinetic energy we get the same equation as above.

In the preceding we have supposed the motion to be vertical; but the result will be unchanged if the motion is not vertical, provided no force except gravity act on the body in the direction of its motion. Any force perpendicular to the motion will do no work and cause no change of kinetic energy. Suppose, for example, the body slides down a smooth plane through a distance s along the plane. Now, we have already shown that it acquires the same velocity as if it fell vertically a distance equal to the height of the plane (§35). Then, if H be the height of the top of the plane and h that of the bottom, the general equation above will still hold. The same is true if the descent is along a smooth curve; for a curve may be regarded as made up of very short straight parts to each of which the principle stated will apply. These results are now readily understood by considering the work performed by gravity. For the total amount of motion in the direction of the whole force of gravity is $(H-h)$. Thus the gain of kinetic energy in the descent from the higher level to the lower must be the same as if the fall were vertical.

60. Kinetic Energy and Elasticity.—When a body is acted on by the force due to a stretched spiral spring, the spring will do work on the body if the spring is contracting, and the body will do work against the force of the spring if it is moving so as to stretch the spring further. Let us first suppose that the body is moving toward the spring with a velocity v, the spring being at that moment stretched to an amount X beyond its normal or unstretched length. While the spring is contracting the velocity of the body will be constantly increasing. Let the velocity be V when the spring has contracted so that its stretch is decreased to x. In this time (§56) the spring will have done an amount of work $(\frac{1}{2}kX^2-\frac{1}{2}kx^2)$ and, since this must equal the increase of kinetic energy of the body,

$$\tfrac{1}{2}kX^2-\tfrac{1}{2}kx^2=\tfrac{1}{2}mV^2-\tfrac{1}{2}mv^2$$

We may also suppose the case reversed, that is, we may suppose the body to be moving away from the spring with a velocity V

when the stretch of the spring is x. Then the velocity of the body will decrease; and, if it be v when the stretch of the spring is X, work $(\frac{1}{2}kX^2-\frac{1}{2}kx^2)$ will have been done against the spring and the decrease of the kinetic energy will be $(\frac{1}{2}mV^2-\frac{1}{2}mv^2)$. Equating these we get the same equation as before.

61. Potential Energy.—We shall now consider the two illustrations just given from another point of view. In the case of a body projected vertically upward, there is a loss of kinetic energy equal to mg multiplied by the height of ascent; and, if the body be allowed to descend again, the same amount of work will be performed by gravity and the body will regain its lost kinetic energy. Thus at the higher level the body (or rather the body and the earth regarded as one system) has an advantage of position that is equivalent to a certain amount of kinetic energy lost, and this advantage of position is measured by $mg(H-h)$. This, since it is equivalent to a certain amount of kinetic energy, is called *potential energy*. Thus it follows that *the sum of the kinetic energy and the potential energy is a constant*, a fact brought out more clearly by writing the equation of §59 thus:

$$\tfrac{1}{2}mV^2+mgh=\tfrac{1}{2}mv^2+mgH$$

Here mgh is the increase of the potential energy when the body is raised from the arbitrary zero level (*e.g.*, sea-level) from which h is measured to the height h, and a similar statement applies to mgH. When the body is at the zero level, it and the earth still possess potential energy, since work could be obtained by allowing the body to fall down a vertical shaft.

Again, in the case of the work done against a spring by a moving body, there is a decrease of kinetic energy, and this decrease is equal to the work done against the spring. If the motion be reversed, the lost kinetic energy will be regained. Thus when the stretch of the spring increases from x to X the spring acquires a capacity for doing work of the amount $(\frac{1}{2}kX^2-\frac{1}{2}kx^2)$, equal to the kinetic energy lost by the body; and the spring yields up this capacity for doing work in restoring the kinetic energy of the body. Writing the equation of §60 in the form

$$\tfrac{1}{2}mV^2+\tfrac{1}{2}kx^2=\tfrac{1}{2}mv^2+\tfrac{1}{2}kX^2,$$

we see that the sum of the kinetic energy of the body and the work the spring can do in contracting to its unstretched length is a constant. The work the spring can do in contracting to its unstretched length is the potential energy of the spring.

In the case of a body separated from the earth the potential energy of the body and the earth depends on their *relative position*, and in the case of the energy of the spring the potential energy depends on the relative positions of the parts of the spring. Hence we may say that *potential energy is the capacity a body or system of bodies has for doing work in virtue of the relative positions of its parts.*

In the case of potential energy we cannot give any universal formula by which it can be calculated as we can in the case of kinetic energy. In each case of potential energy we must calculate how much work the body or system can do in passing from one state to another, and take this as the difference of the potential energy of the body or system in the two states. For any one particular case of potential energy we may deduce a special expression for its amount, such as those given above for gravity and elasticity.

From the statements made in §§58–61, it is evident that we may define *energy*, of either kind, as *the capacity for doing work.*

62. Interchanges of Kinetic and Potential Energy.—We have considered somewhat fully two cases of the interchange of kinetic and potential energy, namely, those of gravity and elasticity, because these are typical and are easily worked out by elementary methods. Such interchanges are common in nature and in industry and a few may be briefly stated.

(*a*) *Change from Kinetic to Potential.*—When a block of wood is split by a wedge or axe, the axe or sledge hammer loses kinetic energy and potential energy of separation of the particles of wood is produced.

As the distance of the earth from the sun increases from midwinter to midsummer, the speed of motion and the kinetic energy decrease and the potential energy of separation increases.

(*b*) *Change from Potential to Kinetic.*—A clockw-eight or watch-spring when wound up has potential energy, and this changes to kinetic energy of the pendulum or balance wheel, which would otherwise come to rest.

A bent bow has potential energy due to the change of position of the particles of the bow and the forces between them. As it unbends it loses this potential energy and the arrow gains kinetic energy.

Water in a lake or above a dam has potential energy; when allowed to escape to a lower level, it loses part of its potential energy and either gains kinetic energy itself or, if it acts on a water wheel or turbine, it imparts kinetic energy to the latter.

(*c*) *Periodic Interchanges.*—In any case of vibration *energy continually changes from the kinetic to the potential form and back again.* Thus, in the vibration of a pendulum, at the bottom of the arc of vibration the potential energy is at a minimum and the kinetic energy is at a maximum, while, at the end of the arc of vibration, the kinetic energy is zero and the potential energy has increased to a maximum. Similar statements apply to the vibration of a tuning fork, a violin string, a body attached to the end of a wire and vibrating torsionally, the oscillations of the balance-wheel of a watch, and so on.

63. Two Kinds of Forces.—In the preceding we have seen that, when the forces acting between bodies are forces of gravity or forces of elasticity, their action leaves the total kinetic and potential energy of the bodies unchanged, or, as it is usually stated, when only such forces act the total kinetic and potential energy is *conserved.* Forces whose action between bodies does not cause a change of the total kinetic and potential energy of the bodies are called *conservative forces*, and any system of bodies between which the forces are wholly conservative is called a *conservative system.*

In contrast with these conservative forces stands such a force as friction. A moving body opposed by friction loses kinetic energy as its velocity decreases, but it does not at the same time gain potential energy to an equivalent extent. Thus, a body started up a rough inclined plane with a certain velocity will not reach as high a level as it would reach if the plane were smooth, and it will not have as much potential energy when it reaches its highest point. Moreover, its descent will be further opposed by friction, and its store of kinetic and potential energy will thereby be further reduced. Friction, then, is a non-con-

servative force since, when in action, it causes a permanent decrease of the kinetic and potential energy of a system.

The reason why such a force as gravity has no effect on the sum total of kinetic and potential energy is easily seen. At a certain distance of a body from the earth the force between the two depends only on their distance apart, and is independent of the way in which they are moving. Hence, when they are moving away from each other and are a certain distance apart, they are losing kinetic energy at a rate exactly equal to the rate at which they regain kinetic energy when, at the same distance of separation, they are moving toward one another. Thus forces of gravity between bodies *depend only on the relative positions of the bodies.* The same is true of the forces between the parts of an elastic spring, and this accounts for the fact that such forces of elasticity are also conservative; in fact it is the fundamental characteristic of all conservative forces. But a non-conservative force, such as friction, depends on the way in which a body or a system of bodies is moving; it is always opposed to the direction of relative motion of bodies in contact; hence it causes a diminution of the kinetic energy of the bodies in whichever direction motion is taking place.

64. The Conservation of Mechanical Energy.—We have seen in the preceding that *under certain conditions the total kinetic and potential energy of a system is constant or is conserved.* The conditions referred to are two, (1) the system must not receive energy from or give energy to any outside bodies, (2) the forces between the parts of the system must be wholly conservative. In reality no system wholly satisfies these conditions. No system is wholly *isolated* in the sense implied in the first condition; and non-conservative forces, such as friction, are never quite absent. But in many cases these conditions are very nearly satisfied. The solar system, consisting of the sun, planets, and moons, is practically isolated; and, while there are internal frictional forces such as those of the tides, the work they do is so small compared with the total energy of the system, that their effects in reducing the kinetic energy of the whole have not yet been detected with certainty. Again, the system consisting of the earth and a body vibrating as a pendulum in a vacuum is practically an isolated system free from frictional forces, and the total kinetic and

potential energy is very nearly constant; the same is true of a heavy body attached to a spring and vibrating in a vacuum. When, as in cases like these, the conditions are sufficiently nearly satisfied, the principle of the constancy of kinetic and potential energy will often lead to valuable results.

In an isolated system in which there are non-conservative forces, such as friction, energy is expended in doing work against these forces; and if, to the sum of the kinetic and potential energy, we add the work done against non-conservative forces, the sum will be constant. But what becomes of the energy so expended? For long it was supposed to be wholly lost. It was, of course, known that heat was produced when work was done against friction; but heat was supposed to be a form of matter. But about 1840 the view was advanced that heat, instead of being a form of matter, is a form of energy as this word is now defined, and this led to the discovery of the **Law of Conservation of Energy**, which is treated fully later under "Heat."

KINEMATICS OF RIGID BODIES

ROTATION

65. Angular Displacements.—In §§9–35, we studied the motion of translation of a point, as a preliminary to the study of the effect of forces on the motion of particles and of bodies moving without rotation. We shall now consider the motion of bodies in rotation, as a preliminary to studying the effects of forces on the motion of rotation of bodies.

The motion of a body is one of rotation when each point in the body moves in a circle the center of which is on a straight line called the axis of rotation. All points in the body turn in any time through equal angles and the angle described in any time is called the *angular displacement* of the body in that time. Its magnitude may be stated in degrees or in radians (1 radian $= 57.3°$ approx.), but the latter method is in many ways the more convenient for the present purposes.

66. Angular Velocity.—The rate of rotation of a body is called its *angular velocity*. When the angular displacements of a body in all equal times are equal, the velocity is a constant angular velocity, and the magnitude of the angular velocity is *the angle*

through which the body turns in unit time. If the angle is reckoned in radians and the second is taken as unit of time, the magnitude of the angular velocity is the number of radians described in one second. The unit of angular velocity is *one radian per second.*

If the velocity is not constant, as, for example, when a fly-wheel is being set in motion or stopped, the angular velocity or rate of angular displacement is defined in the same way as in the analogous case of variable linear velocity (§19), that is to say, we must take the average angular velocity in a short time and then suppose this time indefinitely decreased, so that the average angular velocity approaches a limiting value, which is the *instantaneous angular velocity.*

67. Angular Acceleration.—The rate of increase of the angular velocity of a body is called its *angular acceleration.* When the angular velocity increases by equal amounts in equal times, the angular acceleration is constant and its magnitude is the *increase of angular velocity in unit time.* If we denote the angular acceleration by α, the increase of angular velocity in each second is α and the increase in t seconds is αt Hence, if at the beginning of an interval of time t the angular velocity is ω_0 and at the end of the interval it is ω,

$$\omega = \omega_0 + \alpha t \tag{1}$$

In this time the body has turned through a certain angle say ϕ. To find the magnitude of ϕ we may represent the varying values of the angular velocity by means of a curve of angular velocity, as we did in the similar case of a varying linear velocity (§27), and the area of the diagram will represent the angle ϕ. The two diagrams would have precisely similar properties, the only difference being that in one case we would speak of linear displacement, s, linear velocity, v, linear acceleration, a, while in the other case we would speak of angular displacement, ϕ, angular velocity, ω, and angular acceleration, α. Hence, when the angular acceleration is constant, the formula for ϕ, which must be precisely similar to (2) of §27, is

$$\phi = \omega_0 t + \tfrac{1}{2}\alpha t^2 \tag{2}$$

By elimination of t between (1) and (2) we get

$$\omega^2 = \omega_0^2 + 2\alpha\phi \tag{3}$$

68. Angular Velocity and Linear Velocity.—When a point revolves at a constant rate in a circle, its motion may be described either by means of its angular velocity, ω, or by its linear velocity, v, along the tangent, and there is a simple relation between the two. Let the radius of the circle be r and let the time in which the point moves from P to Q be t. Denoting the length of the arc PQ by s and the angle POQ by ϕ, we have, from the definitions of linear and of angular velocity,

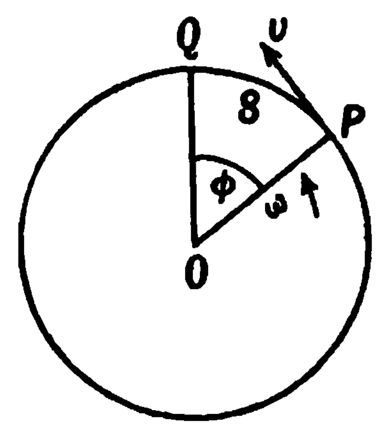


FIG. 31.

$$s = vt. \qquad \phi = \omega t.$$

Now in radian measurement $\phi = s/r$. Substituting in this the above values of s and ϕ, we get

$$\omega = v/r$$

Thus the relation between angular velocity and linear velocity when a point rotates in a circle is the same as the relation between an angle and the arc which it subtends.

The above relation is important. More briefly stated, the proof of it amounts to this: v is the length of arc described per second; hence v/r is the angle described per second in radian measurement, that is the angular velocity.

When a point describes a circle with variable speed, the above relation holds true, with the understanding that ω and v are the instantaneous values of the angular and the linear velocity respectively. The proof is the same as above, t being taken as a very short interval.

When a body rotates about an axis with angular velocity ω, a point in the body describes a circle of radius r, and r is different for points at different distances from the axis. If r and r' are the respective distances of two points from the axis and v and v' their respective linear velocities, $v = r\omega$ and $v' = r'\omega$. Hence $v : v' :: r : r'$.

69. Instantaneous Axes of Rotation.—When the axis about which a body rotates varies from moment to moment, the above relation is true of the values of ω, v, and r at any moment. For example the wheel of a moving wagon or bicycle is always in contact with the road and the point of contact is at any moment

the point about which the whole wheel is at that moment rotating. Now the top of the wheel is twice as far from the ground as the center of the hub and must, therefore, have twice as great a linear velocity.

70. Angular Acceleration and Linear Acceleration.—When a point revolves in a circle (Fig. 32) with changing angular velocity, it has an angular acceleration, say α. The speed of the point along the tangent increases with an acceleration, say a. The same relation holds between a and α as between v and ω (§68). For, if ω is

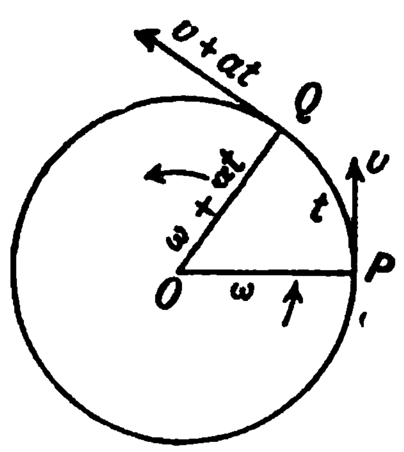


FIG. 32.

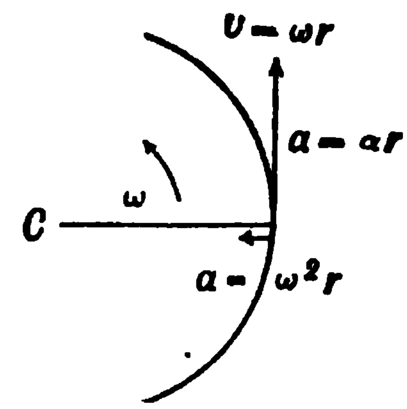


FIG 33.

the angular velocity at the beginning of a short time t and v the linear speed at this time, $v=\omega r$; at the end of the time t the angular velocity is $(\omega+\alpha t)$ and the linear speed is $(v+at)$. Hence $(v+at)=r(\omega+\alpha t)$. Subtracting the former equation from the latter and cancelling t, we have

$$a=r\alpha$$

More briefly stated, a is the added linear speed per unit time and α the added angular velocity per unit time, and the relation between angular velocity and linear speed must hold true of these increases.

It should be carefully noted that a here means the rate of change of speed along the tangent. Since the direction of the velocity is also changing this cannot be the only acceleration. In fact, as we have already seen (§32), there is in all cases of motion in a curve a linear acceleration toward the center equal to v^2/r, or, as we may now write it, $\omega^2 r$, since $v=r\omega$.

The above relations, which are very important, are summarized in Fig. 33.

71. Graphical Representation of Angular Quantities.—An angular displacement is of a certain magnitude and is about a certain

axis. Given the axis, the direction of rotation around it, and the magnitude of the angular displacement, we know everything about it. Now all these can be represented graphically by *a length marked off on the axis* so as to represent to some scale (*e.g.*, a cm. per radian) the magnitude of the angular displacement. There must also be some agreement as to which direction along the axis shall represent a certain direction of rotation about the axis. The rule usually adopted for this purpose is called the "right-handed screw rule," namely, *let the direction* (along *the axis*) *of the line that represents an angular displacement be related to the direction of the rotation as the direction of translation is to the direction of rotation of an ordinary* (*right-handed*) *screw.* For example, a line to represent the angular displacement of the earth in 24 hours due to its rotation about its axis would be drawn from the center toward the N. pole. Two lines to represent the angular displacements of the hands of a watch in one hour would be drawn through the center of the face toward the back and the one for the minute hand would be twelve times as long as the one for the hour hand.

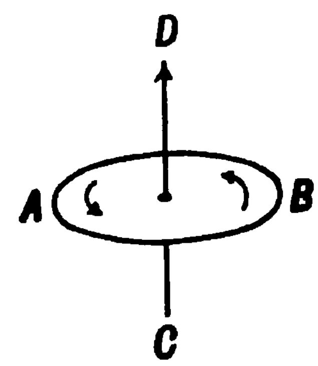


FIG. 34.—A rotation indicated by the arrows is represented by *CD*.

A line to represent an angular velocity would be laid off on the axis of rotation according to the above rule, and a line to represent an angular acceleration would be drawn in the same way.

A directed line that represents according to the above agreement an angular displacement is a vector, since it has both magnitude and direction; but it differs from vectors that represent linear displacements in the fact that it must in any diagram be located on a certain line, namely, the line that stands for the axis of rotation. Such a vector is therefore called a *localised vector* or *rotor*. Two parallel and equal vectors of this kind, not in the same line, do not represent the same angular displacement, since the rotations they represent are about different axes.

72. Addition of Angular Velocities and Accelerations about Intersecting Axes.—A body may have two or more simultaneous angular velocities. For example, suppose a bicycle wheel while rotating about its axis is mounted on a horizontal platform which is kept in rotation about a vertical axis. At any moment the wheel has two component angular velocities about intersecting axes. Each may be represented by a vector drawn from the center according to the rule stated in §71. We may then add

these two vectors by the parallelogram method, and the diagonal will represent in magnitude and direction the resultant angular velocity at the moment in question. (This is fully proven in advanced treatises.)

Since angular accelerations are increments of angular velocities per unit time, we may add them as we add angular velocities.

It follows from the above that angular velocities about *intersecting axes* may be compounded and resolved by the methods applicable to linear velocities (§§23–25), and a similar statement holds true for angular accelerations about intersecting axes.

DYNAMICS OF RIGID BODIES

CENTER OF MASS

73. General Description of Center of Mass.—When the motion of a rigid body is one of translation without rotation, all points in the body move in exactly the same way, and, in describing or calculating the motion, any point in the body may be taken as representing the whole body. When the motion is one of translation combined with rotation, different points in the body move differently and there is no one point the motion of which completely represents the motion of the whole body. There is, however, in any body one particular point which, for many purposes, may be taken as representing the body, so that for these purposes the body may be regarded as concentrated to a particle at that point. This point, which we shall define presently, is the *center of mass* of the body. For instance, let a uniform circular disk be tossed into the air; it will be seen that the center of the disk moves like a particle either in a straight line or in a parabola, while other points in the disk rotate around it. If the disk is loaded with lead on one side it will be some other point, not the geometrical center, that will show this property.

If a body wholly free were struck a blow at random, it would start with both translation and rotation; but if the blow were applied at the center of mass or in a line through the center of mass, the motion would be one of translation without rotation.

The center of mass is thus seen to be a point of great importance in describing or calculating the whole motion of a body. In what follows we shall define the center of mass and show how

its position may be calculated. Then from the definition we shall deduce the above and other properties.

74. Center of Mass of a Number of Particles.—The meaning of the center of mass, in general, will be more clearly understood if we begin with some simple cases that suggest the general definition. (1) *Two Particles.* Let the particles be m_1 at P_1 and m_2 at P_2. Let C_1 be a point that divides P_1P_2 inversely as the masses of the particles, that is, such that

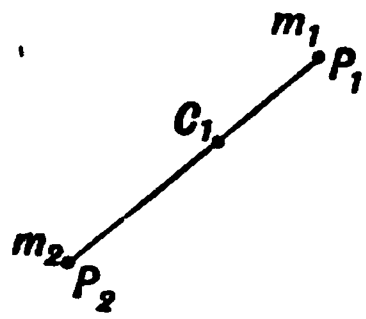


FIG. 35.

$$m_1 \cdot C_1P_1 = m_2 \cdot C_1P_2$$

C_1 is the center of mass of m_1 and m_2.

(2) *Three Particles.* Let the particles be m_1 and m_2 as above and m_3 at P_3. Suppose m_1 and m_2 replaced by (m_1+m_2) at C and let C_2 be a point in C_1P_3 such that (Fig. 36)

$$(m_1+m_2)C_1C_2 = m_3 \cdot C_2P_3$$

C_2 is the center of mass of m_1, m_2 and m_3.

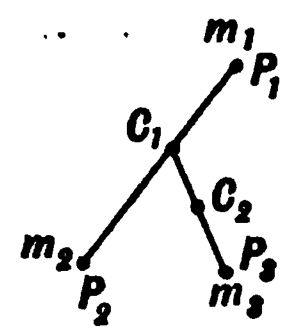


FIG. 36.

(3) *Any Number of Particles.* Proceeding as above we get the center of mass, C, of any number of particles, and the same will apply to a body of any form, since it may be divided up into a large number of small parts.

We shall show in the next section that the point to which such a process leads is independent of the order in which the particles are taken.

75. Distance of Center of Mass from a Plane.—Let EF (Fig. 37) be the line in which any plane is cut by a perpendicular

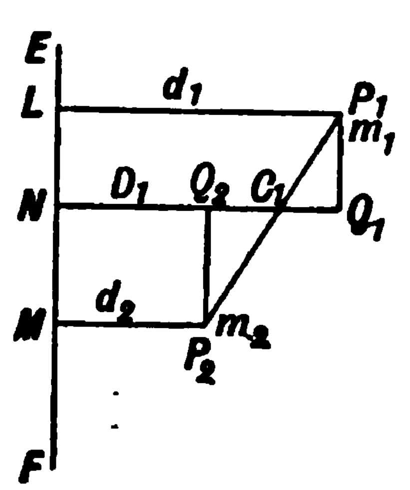


FIG. 37.

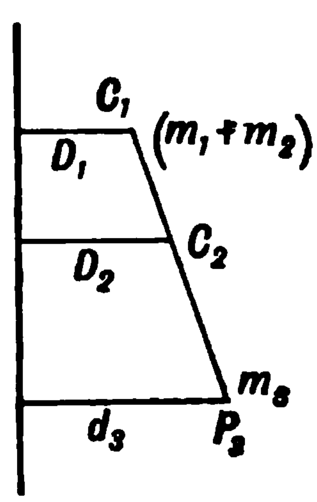


FIG. 38.

plane through P_1P_2 of the last section. Draw P_1L, P_2M, C_1N perpendicular to EF and denote their lengths by d_1, d_2 and D_1,

respectively. Draw P_1Q_1 and P_2Q_2 perpendicular to C_1N. Since C_1Q_1 and C_2Q_2 are the projections of C_1P_1 and C_1P_2 it is readily seen from the equation in §74 (1) that

$$m_1 \cdot C_1Q_1 = m_2 \cdot C_1Q_2$$

or
$$m_1(d_1 - D_1) = m_2(D_1 - d_2)$$
$$\therefore (m_1 + m_2)D_1 = m_1d_1 + m_2d_2$$

If we should proceed to apply the same method to $(m_1 + m_2)$ at C_1 and m_3 at P_3 (Fig. 38) we would, it is evident, get a similar result. Hence

$$(\overline{m_1 + m_2} + m_3)D_2 = (m_1 + m_2)D_1 + m_3d_3$$

Hence, substituting from the above,

$$(m_1 + m_2 + m_3)D_2 = m_1d_1 + m_2d_2 + m_3d_3$$

By extending the same method to any number of particles we shall evidently obtain the general formula

$$(m_1 + m_2 + \cdots)D = m_1d_1 + m_2d_2 + \cdots$$

It is evident that this result will not be altered if the order in which the various particles are taken be altered in any way, and that it is true whatever plane of reference is chosen.

76. General Definition of Center of Mass.—If $m_1, m_2, \cdots$ are the respective masses of the particles constituting a body (or group of particles) of total mass M, and if the respective distances of these particles from any plane are $d_1, d_2, \cdots$, the center of mass is a point whose distance from the plane is

$$D = \frac{m_1d_1 + m_2d_2 + \cdots}{m_1 + m_2 + \ldots} = \frac{\Sigma md}{M}$$

If in any case one or more of the distances are measured on the opposite side of the plane from the others, when we substitute numbers for the various distances those corresponding to one side of the plane must be given positive signs and the others negative.

If the plane from which $d_1, d_2 \cdots$ are measured passes through the center of mass, D is zero and in this case

$$m_1d_1 + m_2d_2 + \cdots = 0$$

When in any case it is desired to find the position of the center of mass of a body by applying the above formula, it is only

necessary to apply it to distances from three planes at right angles. Denoting the distances from one of them by x's, from a second by y's, from the third by z's, we get

$$\bar{x}=\frac{\Sigma mx}{M}, \quad \bar{y}=\frac{\Sigma my}{M}, \quad \bar{z}=\frac{\Sigma mz}{M}$$

where $\bar{x}$ denotes the distance of the center of mass from the plane from which the x's are measured, and similarly for $\bar{y}$ and $\bar{z}$.

77. Center of Mass of a Regular Body.—The center of mass of two equal particles is at the middle of the line joining them. A uniform rod may be divided into pairs of equal particles, the two in each pair being equidistant from the center of the rod. Hence the center of mass of the whole rod is at its middle point. Similar reasoning may be applied to any homogeneous body which has a geometrical center such as a circle, ellipse, sphere, spheroid, parallelogram, cube, parallelopiped, etc. The center of mass of each of these is at its geometrical center.

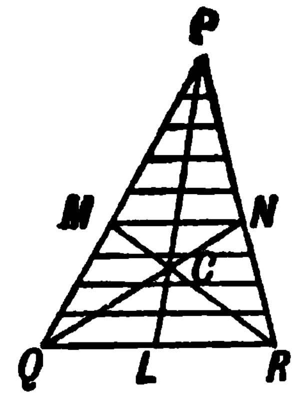


FIG. 39.

When a body can be divided into parts such that the center of mass of each is known, the center of mass of the whole can usually be found. A triangle may be divided into narrow strips parallel to one side; the center of mass of each strip lies on the line joining the middle of that side to the opposite vertex. Hence the center of mass of a triangle is at the intersection of the three lines which join the vertices to the middle of the opposite sides. Similar reasoning shows that the center of mass of a triangular pyramid is at the intersection of the four lines that join the vertices to the respective centers of mass of the opposite faces.

78. Velocity and Acceleration of the Center of Mass.—Let us suppose that the velocity of each particle in a group of particles is known. How can the velocity of the center of mass be found? To answer this it is sufficient to show how the velocity of the center of mass in a direction perpendicular to each of three planes at right angles can be found.

To find the velocity of the center of mass in a direction perpendicular to any plane, consider the distances of the particles and of the center of mass from that plane. These are connected by the equation (§75)

$$(m_1+m_2+\cdots)D=m_1d_1+m_2d_2+\cdots \tag{1}$$

At a time t later these distances will have all changed. Let the new values of the distances be d_1', $d_2' \cdots D'$. Then

$$(m_1+m_2+\cdots)D'=m_1d_1'+m_2d_2'+\cdots \tag{2}$$

Subtract each side of (1) from the corresponding side of (2); divide through by t and suppose t decreased without limit. Then $(D'-D)/t$ will become the velocity, say $\bar{v}$, of the center of mass; $(d_1'-d_1)/t$ will become the velocity, say v_1, of m_1 and so on. Hence

$$(m_1+m_2+\cdots)\bar{v}=m_1v_1+m_2v_2+\cdots \tag{3}$$

Thus the velocity of the center of mass is related to the velocities of the separate particles as the distance of the center of mass from any plane is related to the distances of the particles from that plane.

We may now proceed to apply the same reasoning to find the acceleration of the center of mass. Starting with (3) above, let us consider what (3) becomes at a short time t later. We shall thus get two equations. Subtracting one from the other as before, dividing by t, and then supposing t indefinitely short, we get

$$(m_1+m_2+\cdots)\bar{a}=m_1a_1+m_2a_2+\cdots \tag{4}$$

Equation (3) is readily obtained by differentiating (1) with reference to the time (see §19) and (4) is obtained by differentiating (3) (see §31).

79. Acceleration of Center of Mass due to External Forces.— Equation (4) of the last section has a very important interpretation. The term m_1a_1 is, by the Second Law of Motion, equal to the force that acts on m_1 in the direction in which a is measured, which, of course, may be any direction, and similarly for the other particles. Now the forces may be divided into two groups, (1) forces applied from the outside or *external* forces such as gravity acting on the body, pressures and pulls applied to the surface of the body and so on; (2) forces that the particles exert on one another, that is *internal forces*, actions and reactions between the particles. By the Third Law of Motion these internal forces occur in pairs of equal and opposite forces, and the sum of the components of all of them in any direction is zero.

Hence the right hand side of (4) stands for the sum of the com-

ponents, in the direction considered, of all the external forces. Thus if M be the whole mass of the body or group of particles,

$$\bar{a} = \frac{\textit{sum of components of external forces}}{M}$$

Now by the Second Law of Motion this is the expression we would arrive at if we asked, "what acceleration would the center of mass of the body receive if the whole mass were concentrated there and all the external forces were transferred parallel to themselves so as to act at that point?"

Hence *the center of mass of a body moves as if the whole mass were concentrated at the center of mass and the forces acting on the body were transferred, with their directions unchanged, to the center of mass.*

We now see the explanation of the facts stated in §73. In the case of a body tossed into the air gravity is the only external force, and the center of mass moves as if all the mass and weight were concentrated there, that is, it moves as a particle would. Even when a body has its form changed very abruptly by the action of internal forces, as in the case of the explosion of a rocket, the internal forces do not affect the motion of the center of mass of all the particles. When two bodies approach and impinge, the motion of their center of mass is not affected by the forces between the bodies during impact, and hence continues unchanged after the impact. There are powerful forces of attraction between the sun and the planets that make up our solar system, but the center of mass of the whole moves with a uniform velocity through space.

80. Translation and Rotation.—It is evident that to ascertain the whole motion of a body it is sufficient to find: (1) The motion of translation of some point in the body. (2) The motion of rotation about that point. By the result obtained in §79 we can find the linear acceleration of the center of mass from the magnitudes and directions of the forces, without considering the distances of their lines of action from the center of mass.

We shall now consider how the angular acceleration of a body can be calculated, but in an elementary work it is necessary to confine attention, for the most part, to the rotation of rigid bodies mounted on fixed axes.

MOMENTS OF FORCE AND MOMENTS OF INERTIA

81. When a rigid body is mounted on a fixed axis (*e.g.* a grindstone or fly-wheel) the only motion that a force applied to it can produce is one of rotation about the axis. To find the magnitude of the effect we must consider, not only the magnitude and the direction of the force, but also the distance of its line of action from the axis. For it is a matter of common experience that a force can be most effectively applied to set a large body into rotation when it is applied as far from the axis as possible.

On the other hand, the inertia resistance which the force encounters depends on something more than the mass of the body. For it is also well known that the farther, on the whole, the mass of the body is from the axis, as, for example, in the case of a fly-wheel with a heavy rim and light spokes, the harder it is to set it into rotation or to stop it.

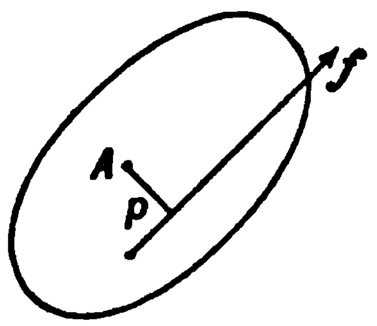


FIG. 40.

We are thus led to consider moments of force and moments of inertia.

82. The Moment of a Force.—Consider a body mounted on a fixed axis, A, perpendicular to the plane of the paper. Let a force, f, act on the body, the line of action of the force being in a plane perpendicular to the axis, and let the perpendicular distance of the line of action from the axis be p. The product fp is called the moment of f about A. It depends on the magnitude, direction, and line of action of the force; but it does not depend on the particular point in the line of action at which f is applied. A moment of force is also called a *torque*.

The above is not a general definition of the moment of a force, for we have supposed the line of action of the force to be in a plane perpendicular to the axis. To find the moment of a force, F, in any direction, we must first resolve F into a component parallel to the axis and a component, say f, perpendicular to the axis. The former cannot produce motion about the axis, since it is parallel to the axis. The latter component, f, is the effective component.

The moment of a force about an axis is the product of the component of the force perpendicular to the axis (the other component

being parallel to the axis) by the perpendicular distance of this component from the axis.

Since one direction of rotation about an axis is taken as positive, the other being taken as negative, moments of forces are considered as positive or negative according to the directions in which they tend to produce rotation.

A moment of force, although it is the product of two quantities f and p, should be thought of as a single physical quantity, just as work, the product of F and s, is a single physical quantity. As such we shall denote it by L.

83. Work Done by a Moment of Force.—Let the line of action of the force be fixed relatively to the body, so that the moment of force about the axis is constant and equal to fp. If the body rotate through an angle θ in the direction of the moment, the force will have acted through $p\theta$ and the work done will be $fp\theta$ or $L\theta$. Hence in rotation,

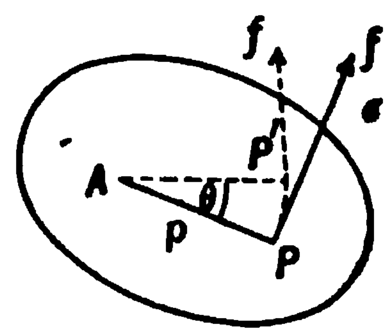


FIG. 41.

Work = moment of force × angular displacement

The similarity of this expression to that for the work done in translation, namely Fs, should be noted, moment of force corresponding to force and angular displacement to linear displacement.

84. Kinetic Energy of Rotation.—Each particle of a rotating body has a certain linear velocity and a certain amount of kinetic energy, and the total kinetic energy is the sum of the kinetic energies of the particles. A particle of mass m at a distance r from the axis has a linear velocity ωr where ω is the angular velocity, and its kinetic energy (in absolute units) is $\frac{1}{2}m\omega^2r^2$. Now r is different for different particles but ω is the same for all.

Hence

$$E = \tfrac{1}{2}(m_1r_1^2 + m_2r_2^2 + \ldots)\omega^2$$

The term in brackets evidently depends on the mass and form of the body and on the particular axis of rotation considered. If we denote it by I,

$$E = \tfrac{1}{2}I\omega^2$$

an expression similar to that for kinetic energy of translation, namely $\frac{1}{2}mv^2$.

85. Moments of Inertia.—The expression denoted above by I is called the *moment of inertia* of the body about the particular axis of rotation. It may be defined as *the sum of the products of the particles by the squares of their respective distances from the axis of rotation*, or briefly,

$$I = \Sigma mr^2.$$

The value of this sum, in the case of a body of regular geometrical shape, can be expressed in terms of its mass and its linear dimensions. For example, all parts of a thin hoop of mass M and radius r may be regarded as being at the same distance from its geometrical axis and its moment of inertia about that axis is, therefore, Mr^2. Two other important cases are the following: For a circular disk about its geometrical axis (Fig. 42).

$$I = \tfrac{1}{2}Mr^2$$

For a thin red about a transverse axis through its center (Fig. 43).

$$I = \tfrac{1}{12}Ml^2$$

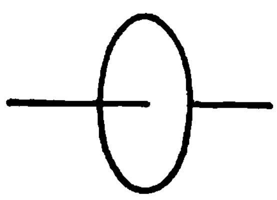


FIG. 42.

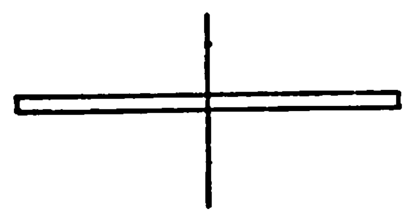


FIG. 43.

The derivation of such formulæ is best performed by means of the Integral Calculus.

If ρ be the mass per unit area of the disk, and if it be supposed divided into hoops, I is the integral from 0 to r of $2\pi r dr \rho r^2$ which is $\frac{1}{2}\pi r\rho^4$ or $\frac{1}{2}Mr^2$.

If ρ be the mass of unit of length of the rod, I is the integral from $-l/2$ to $l/2$ of $\rho dr r^2$ which is $\frac{1}{12}\rho l^3$ or $\frac{1}{12}Ml^2$.

If the moment of inertia of a body about a certain axis is I and the mass of the body is M, and if we take a length k such that $I = Mk^2$, k is called the radius of gyration of the body about that axis.

86. Energy Equation for Rotation.—Consider a rigid body on a fixed axis and acted on by a moment of force L, the moment of inertia about the axis being I. If the body turn through an angle θ the work done will be $L\theta$. Since the body is rigid, the relative positions of its particles will not be changed and there will, therefore, be no change of potential energy. Hence the work done will equal the increase of kinetic energy, or

$$L\theta = \tfrac{1}{2}I(\omega^2 - \omega_0^2)$$

87. Angular Momentum.—In the case just considered let the time of rotation through the angle θ be t. The average angular velocity in the time t is $\frac{1}{2}(\omega+\omega_0)$. Hence

$$\theta=\frac{1}{2}(\omega+\omega_0)t$$

From this and the equation of §86 we get

$$Lt=I(\omega-\omega_0)$$

The product Lt evidently corresponds to the product Ft in the case of translation (§45) and may be called the impulse of the moment of force. The expression on the right is the increase of $I\omega$. From the analogy of momentum, mv, the *product of moment of inertia and angular velocity* is called *angular momentum.*

If $L=0$, that is, if the body is not acted on by any force having a moment about the axis of rotation, $I\omega=I\omega_0$, or the angular momentum is constant.

88. Conservation of Angular Momentum.—

The last statement is a particular case of the principle called the Conservation of Angular Momentum, namely, the case of a rigid body mounted on a fixed axis. A wobbling quoit is a more general case. Neglecting air resistance, the only external force is gravity, which acts through the center of mass. The angular momentum is constant in amount but its axis has a periodic motion. An acrobat turning in the air and the projectile from a rifled gun are other illustrations. The whole solar system illustrates the general principle, which is that the total angular momentum of any system of bodies not acted on by forces having a resultant moment about the center of mass is constant in amount.

89. Angular Acceleration Produced by a Moment of Force.—From §87 it follows that

$$L=I\frac{\omega-\omega_0}{t}$$

Hence if α be the angular acceleration produced by L

$$L=I\alpha$$

This is the fundamental equation for calculating the motion of a body on a fixed axis. It corresponds to $F=ma$ for translation.

If the body is not on a fixed axis but is free, the applied force will produce linear acceleration of the center of mass (§79) and also rotation about some axis through the center of mass, the direction of which depends on the shape of the body. If the line of action of the force is in the plane of two of

the principal axes (§101), the rotation will be about the third, but the subject cannot be discussed fully here.

90. Moments of Inertia about Parallel Axes.—There is a simple and useful relation between the moment of inertia, I, of a body about any axis and its moment of inertia, I_0, about a parallel axis through the center of mass, namely,

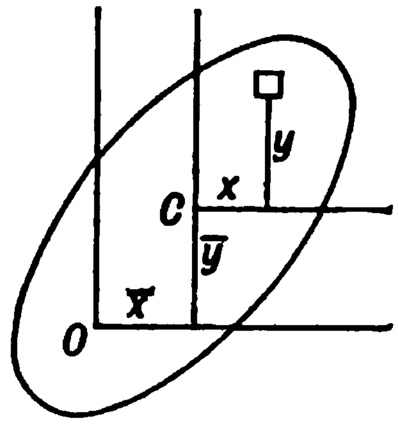


Fig. 44.

$$I = I_0 + Mh^2,$$

M being the mass of the body and h the distance between the two axes. The proof of this is as follows:

Let the axes referred to be perpendicular to the plane of Fig. 44 and cut the plane in O and C respectively. Let the coördinates of a particle, m, referred to rectangular axes through C be x and y and let the coördinates of C referred to a parallel set of axes through O be $\bar{x}$ and $\bar{y}$. Then

$$I = \Sigma m\{(x+\bar{x})^2 + (y+\bar{y})^2\}$$
$$I_0 + Mh^2 = \Sigma m(x^2+y^2) + \Sigma m(\bar{x}^2+\bar{y}^2)$$

The difference between the right hand sides is zero, since x and y are distances from planes through the center of mass and (§76) $\Sigma mx = 0$, $\Sigma my = 0$.

91. Kinetic Energy of a Body which has Translation and Rotation.—Let the body of Fig. 44 be in rotation about the axis through O, supposed fixed in the body (or in fixed connection with it) with angular velocity ω. This is also its angular velocity about the axis through C, since both revolutions are completed in the same time. The total kinetic energy is $\frac{1}{2}I\omega^2$. From the equation of §90,

$$\tfrac{1}{2}I\omega^2 = \tfrac{1}{2}I_0\omega^2 + \tfrac{1}{2}Mh^2\omega^2$$
$$= \tfrac{1}{2}I_0\omega^2 + \tfrac{1}{2}M\bar{V}^2$$

since $\bar{V}$, the linear velocity of C, is equal to $h\omega$. Thus the total kinetic energy may be regarded as consisting of

(1) K. E. of Translation of the C. of M.
(2) K. E. of Rotation about the C. of M.

This applies to a cylinder rolling down a plane or a moving carriage wheel, since there is an axis which is, for the moment, at rest,

namely, the line of contact with the surface. It holds, likewise, for any body rotating about an axis through the center of mass and also moving perpendicular to that axis, for there is in all such cases an instantaneous axis (§69). In fact, by resolving the linear velocity parallel to and perpendicular to the axis, it will be seen that the principle is true for any motion of a rigid body.

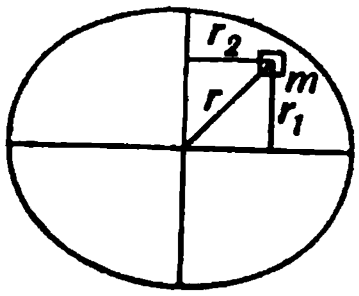


FIG. 45.

92. Moments of Inertia of a Disk.—If the moment of inertia of a disk of any shape about two axes at right angles in the plane of the disk be I_1 and I_2, its moment of inertia about a third axis intersecting the two and perpendicular to the plane of the disk is I_1+I_2.

For let the distances of an element, m, of the disk from the first two axes be r_1 and r_2 respectively Fig. 45.

$$r^2=r_1^2+r_2^2$$
$$mr^2=mr_1^2+mr_2^2$$

Summing up for all elements

$$I=I_1+I_2.$$

93. Some other Moments of Inertia.—

A uniform rectangular disk may be divided into rods and their moments of inertia added. Thus about axes in the plane of the disk and bisecting pairs of opposite sides $I_a=\frac{1}{12}Ma^2$, $I_b=\frac{1}{12}Mb^2$. Hence the moment of inertia of the disk about an axis through the center perpendicular to the disk is

$$I=\tfrac{1}{12}M(a^2+b^2)$$

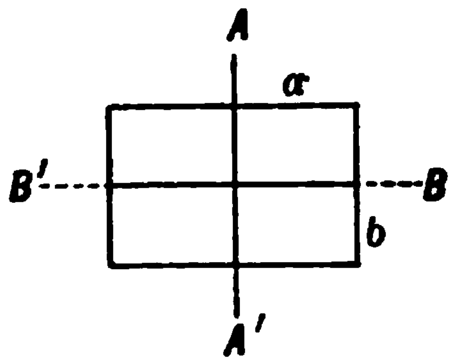


FIG. 46.

A uniform rectangular block may be divided into rectangular disks. Hence its moment of inertia about an axis through its center and perpendicular to a face with sides equal to a and b is given by the last formula.

The moment of inertia of a uniform circular disk about any two axes at right angles in the plane of the disk are equal. Hence by §92, each is equal to $\frac{1}{4}Mr^2$.

A circular cylinder may be divided into circular disks. Hence its moment of inertia about its geometrical axis is $\frac{1}{2}Mr^2$. Its moment of inertia about a transverse axis through its center may also be found by the above principles, but we leave it as an exercise.

94. Gravitational Units.—All of the preceding formulæ are in absolute units. To adapt them to gravitational units, it is only necessary to replace m and M, wherever they occur by m/g and M/g. For at the outset (§84) we took $\frac{1}{2}mv^2$ as the kinetic energy of m, whereas in gravitational units it would be $\frac{1}{2}\,\frac{m}{g}\,v^2$ (Engineers write W instead of m). All formulæ in which mass does not occur explicitly are the same in both systems.

TABLE OF MOMENTS OF INERTIA

Body	Axis	Moment of Inertia
Rod	transverse through end	$\frac{1}{3}Ml^2$
Rod	transverse through middle	$\frac{1}{12}ML^2$
Circular disk	perpendicular through center	$\frac{1}{2}MR^2$
Circular cylinder	longitudinal through center	$\frac{1}{2}MR^2$
Circular cylinder	transverse through center	$M(\frac{1}{4}R^2+\frac{1}{12}l^2)$
Rectangular block	through center perpendicular to face with sides a and b in length	$\frac{1}{12}M(a^2+b^2)$
Sphere	through center	$\frac{2}{5}MR^2$

RESULTANT OF FORCES ACTING ON A BODY

95. Resultant.—When treating of the forces acting on a *particle*, we found that they could always be replaced by a single equivalent force called their resultant. When a number of forces act on a *body*, they are in certain cases equivalent in their effects to a single force, which is called their resultant. As we shall see later, there are other cases in which this is not so.

96. Conditions to be Satisfied by Resultant.—1. The resultant must be competent to produce the actual linear acceleration of the center of mass C, and, therefore, its component in any direction must equal the sum of the components of the acting forces in that direction. This condition is simplified by considering that any actual acceleration of C is made up of three independent components along axes at right angles. Hence *the resultant must have a component in each of three rectangular directions equal to the sum of the components of the forces in these directions.*

2. The resultant must be competent to produce the actual an-

gular acceleration about any axis, and, therefore, its moment about any axis must equal the sum of the moments of the acting forces about that axis. It is, however, not necessary to consider all axes; for the angular acceleration about any axis can be resolved into rectangular components. Hence the second condition is that *the moment of the resultant about each of any set of rectangular axes must equal the sum of the moments of the forces about that axis.*

If a force satisfies the above conditions it is the resultant. We shall now apply these tests to find the resultant of the forces acting on a body in some cases of importance.

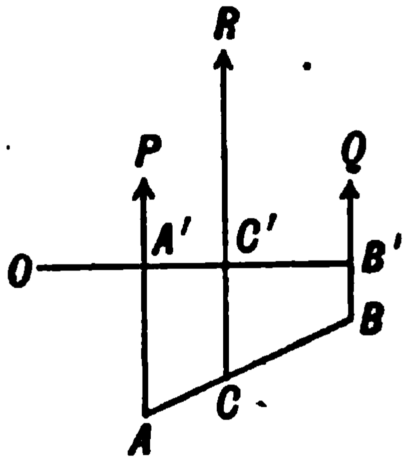


FIG. 47.

97. Resultant of Two Parallel Forces.—1. *Let P and Q be two forces in the same direction* acting at points A and B of a body. A single force R in the direction of P and Q and equal to $(P+Q)$ will satisfy the first condition of §96, since its component in any direction equals the sum of the components of P and Q in that direction.

This force will also satisfy the second condition, provided it acts at a point C in AB such that

$$P \cdot CA = Q \cdot CB$$

For first consider any axis perpendicular to the plane of P and Q. Suppose it to cut that plane in O. Draw $OA'C'B'$ to cut the lines of the forces at right angles. Since $C'A'$ and $C'B'$ are the projections of CA and CB respectively it is readily seen from the last equation that

$$P \cdot C'A' = Q \cdot C'B'$$

or

$$P(OC' - OA') = Q(OB' - OC')$$

$$\therefore (P+Q)OC' = P \cdot OA' + Q \cdot OB'$$

Thus the moment of R about the axis equals the sum of the moments of P and Q. Next take an axis perpendicular to the above axis and to the lines of the forces. All the forces are at the same distance from this axis, and, since R equals $(P+Q)$, the moment of R about it equals the sum of the moments of P and Q about it. Finally an axis perpendicular to the other two will be parallel to P, Q, and R and each will have zero moment about it.

Hence R is the resultant of P and Q.

It is important to notice that C is the point we would have found if we had been seeking the center of mass of particles at A and B proportional to P and Q (§74).

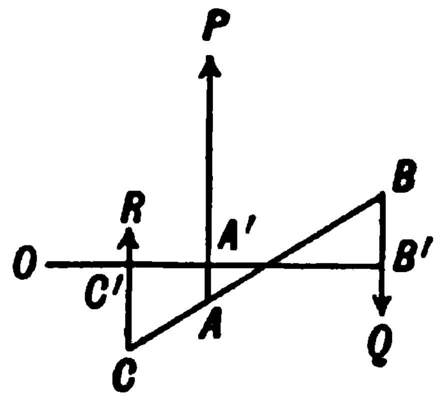


FIG. 48.

2. *Let* P *and* Q *be in opposite directions* and suppose $P > Q$. A single force R in the direction of the greater, P, and equal to $(P-Q)$ will satisfy the first condition, since its component in any direction equals the (algebraic) sum of the components of P and Q in that direction. It will also satisfy the second condition if it acts at a point C in BA produced such that

$$P \cdot CA = Q \cdot CB$$

For consider first an axis perpendicular to the plane of the forces and cutting that plane in O, and draw $OC'A'B'$ to cut the lines of the forces at right angles. Then

$$P \cdot C'A' = Q \cdot C'B'$$

or

$$P(OA' - OC') = Q(OB' - OC')$$

$$\therefore (P-Q)OC' = P \cdot OA' - Q \cdot OB'.$$

Hence the moment of R about the axis equals the (algebraic) sum of the moments of P and Q. The same is true of two other axes taken as in case (1); the proof need not be repeated as it is identical with that there given.

Hence R is the resultant of P and Q.

To find the distance of C from A replace CB by $(CA + AB)$ in the above equation for the position of C. Then

$$CA = \frac{Q \cdot AB}{P-Q}$$

Hence CA is greater the less $(P-Q)$, that is, the more nearly the forces are equal.

98. Couples.—Two equal and opposite forces, not in the same line, constitute a *couple*. If we attempted to find the resultant of two such forces by the method of the last section, it would give zero force at an infinite distance and such a force has no real existence. Hence a couple cannot be reduced to a single force.

The sum of the moments of two forces constituting a couple is the same about all axes perpendicular to the plane of the couple. For, about an axis O between the forces, the moments of the forces are in the same direction and their sum is $(P \cdot OA + P \cdot OB)$ or $P \cdot AB$; and, about an axis O' not between the forces, the moments are in opposite directions and the sum is $(P \cdot O'B' - P \cdot O'A')$ which again equals $P \cdot AB$.

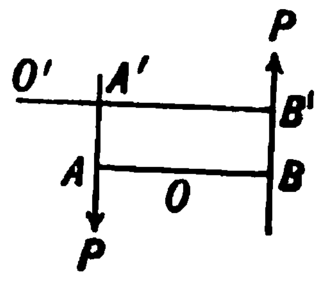


FIG. 49.

The distance AB between the forces of a couple is sometimes called the *arm* of a couple, and the moment of the couple about any axis perpendicular to its plane, that is $P \cdot AB$, is sometimes called the *strength* of the couple. Two couples in the same or parallel planes and of the same strength are equal in all respects and produce equal effects.

Since the sum of the forces of a couple equals zero, the couple produces no acceleration of the center of mass (§79); and if the center of mass be at rest it will remain at rest, or if it be moving in any way it will continue moving with constant velocity. The angular velocity produced by the couple must therefore be about some axis through the center of mass.

99. Resultant of any Number of Parallel Forces.—To find the resultant of any number of parallel forces, whether in one plane or not, we may find the resultant of two, then combine this resultant with a third, and so on. The final resultant will be either a single force or a couple or zero. At each step the resultant equals the algebraic sum of the forces added. Hence the final resultant equals the algebraic sum of all the forces.

The line of action of the resultant may also be found by applying the principle that the moment of the resultant about any axis must equal the sum of the moments of the forces about that axis. When the forces are all in one plane, to find the line of action of the resultant we only need to take moments about any axis perpendicular to the plane. When the forces are not all in one plane, it will be necessary to take moments about two rectangular axes perpendicular to the forces.

100. Center of Gravity.—Attention has been called in (1) §97 to the identity of the method of finding the resultant of parallel forces in the same direction and the method of finding the center

of mass of a number of particles. If, for the particles in a certain group of particles or of a body, we substitute parallel forces all in one direction acting at the respective positions of the particles and proportional to the masses of the particles, the point of action of the resultant will coincide with the center of mass. This is sometimes taken as the definition of the center of mass. It should be noticed that nothing need be said as to the common direction of the parallel forces.

The forces of gravity on the particles of a body are (very nearly) parallel forces and they are proportional to the masses of the particles. Hence the *Center of Gravity* of a body, or the point of action of the resultant of the (very nearly) parallel forces of gravity, coincides with the center of mass of the body.

A very large body near the earth has a definite center of mass but not a definite center of gravity (except in some particular cases), for the forces are not quite parallel nor quite proportional to the masses. This is of no practical importance as regards bodies of the size found on the earth's surface; but it is of great importance in considering the effect of the attraction of the sun and moon on the motion of the earth.

101. Centrifugal Force.—In §47 we found an expression for the force required to keep a particle revolving in a circle. We may now extend this to *a body of any size or shape.* When a body of mass m rotates with constant angular velocity about any axis not through the center of mass, the latter moves uniformly in a circle and has therefore an acceleration v^2/r toward the center. Hence the force acting on the body (or the resultant of the forces if there are several) must, by the principle stated in §79, equal mv^2/r and must act in the line joining the center of mass to the center of the circle, and the body will react with an equal and opposite force. This reaction is the cause of the varying force which an unbalanced fly-wheel exerts on the axis.

In many cases more than a single force (in addition to those required to overcome friction and sustain the weight of the body) is required to keep a body rotating about an axis. As a simple case consider a pair of equal spheres joined by a light rod and rotating about a vertical axis through the center of the rod. Since the center of mass has no acceleration, the forces acting on the body if transferred to the center of mass would have a zero resultant. Hence the forces must form a couple and the reactions on the axis will form a couple, called a *centrifugal couple,* tending to bend the axis or make it rotate about an axis perpendicular to itself. For certain axes of

rotation of a body the centrifugal couple is zero. In the above simple illustration this is true when the axis of rotation is in the line of the centers of the balls or at right angles thereto. These are also the positions of maximum and minimum moments of inertia of the body. A similar statement will evidently apply to a symmetrical body, such as a circular disk, which can be divided into pairs of particles like the above. Whatever the shape of a body there are three rectangular axes through any point of the body about which it can rotate without exerting any centrifugal couple. These are the axis of maximum moment of inertia through the point, that of minimum moment of inertia and a third perpendicular to both. These are called the *principal axes* through the point.

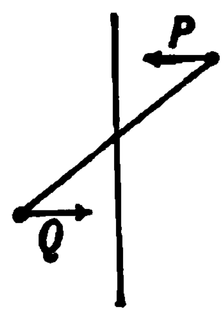


FIG. 50.

When a body is set spinning about a principal axis through its center of mass it continues to spin without any tendency to "wobble" or exert a centrifugal couple. This is illustrated by the motion of a well-known quoit or discus, by that of a bullet from a rifled gun and by the motion of the earth about its axis. But when the axis of initial spin is not a principal axis irregular motion ensues, as is illustrated by a badly thrown quoit.

FORCES IN EQUILIBRIUM

102. Conditions of Equilibrium.—The forces acting on a body are in equilibrium when they cause no acceleration either linear or angular, that is when their resultant is zero.

Given that a system of forces is in equilibrium we may conclude (from §79) that *the sum of their components in any direction equals zero,* since there is no acceleration of the center of mass, and also that *the sum of their moments about any axis equals zero,* since there is no angular acceleration about any axis.

When we equate the sum of the components of the forces in any direction to *zero* we get a relation between the forces, and it might seem that we could get an unlimited number of such relations; but, in reality, there are only three of these independent, *e.g.*, those got by taking the sum of the components in some three directions at right angles.

Similarly we get a relation between the forces by equating the sum of the moments about any axis to zero; but again there are only three of these relations independent, *e.g.*, those got by taking moments about some three rectangular axes.

Thus we can deduce at most six independent relations between forces in equilibrium, and this might have been expected from the fact that a rigid body has six degrees of freedom at most—three of translation and three of rotation.

We may reverse the point of view and ask what relations and how many must forces satisfy to make it certain that they shall be in equilibrium, that is, what are the conditions essential to equilibrium. The answer is again six relations, namely, the sum of the components in each of any three rectangular directions must equal zero and the sum of the moments about each of some three rectangular axes must equal zero.

The two conditions may be stated thus:

$$\Sigma F = 0$$
$$\Sigma Fp = 0$$

for each of any three rectangular axes.

103. Forces in a Plane.—When the lines of action of forces that are in equilibrium lie in one plane, the sum of the components of the forces in each of any two directions at right angles in the plane equals zero. In this case the third rectangular axis is perpendicular to the plane and the component of each force in that direction is zero. Also the sum of the moments of the forces about any axis perpendicular to the plane is zero. The other two rectangular axes are in the plane and the moment of any one of the forces about such an axis is zero.

Hence when forces in a plane are in equilibrium three independent relations among the forces can be deduced.

104. Examples of Equilibrium of Forces in a Plane.—To illustrate the above we shall consider two examples.

W

Fig. 51.

1. A uniform beam AB (length $= l$) rests without slipping on the ground and leans without friction against a smooth wall. What is the force (F_1) at the wall and the vertical force at the ground (F_2) and what is the force of friction (F_3) between the beam and the ground (Fig. 51)?

Since there is no friction at B, F_1 is horizontal. The force of friction at A, that is F_3, is horizontal and toward B. Equating the sum of the horizontal forces acting *on the beam* to zero we get

$$F_1 - F_3 = 0. \qquad (1)$$

and equating the vertical forces to zero we get

$$F_2 - W = 0 \qquad (2)$$

A third relation may be obtained by taking moments about any axis perpendicular to the plane of the forces. If we choose for this purpose an axis through A, the relation will be as simple as possible, since F_2 and F_3 have

zero moment about such an axis. The weight acts at the center C of the beam and the distance of its line of action from A is $(l/2)\cos\theta$. Also the distance BE of the line of action of F_1 from A equals $l \sin\theta$. Hence

$$W\frac{l}{2}\cos\theta - F_1 l\sin\theta = 0 \tag{3}$$

From these three equations we get

$$F_1 = F_3 = \tfrac{1}{2}W\cot\theta$$
$$F_2 = W$$

2. A uniform rod hangs from a wall by a hinge and rests on a smooth floor (Fig. 52). In this case the force at A must be vertical, since there is no horizontal force of friction at A. Let the force on the beam at B consist of a horizontal part F_1 and a vertical part F_3. Equating to zero the sum of the vertical forces, the sum of the horizontal forces and the sum of the moments about B, we get

$$F_1 = 0;\ F_2 + F_3 - W = 0$$
$$W\frac{l}{2}\cos\theta - F_2 l\cos\theta = 0$$

Hence

$$F_1 = 0;\ F_2 = \tfrac{1}{2}W;\ F_3 = \tfrac{1}{2}W$$

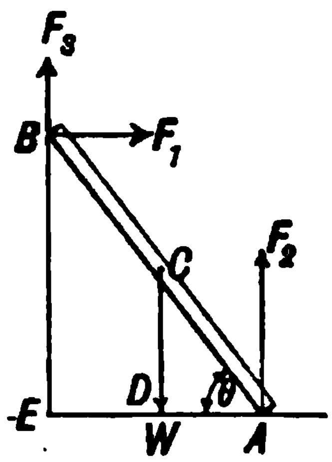


Fig. 52.

Since F_1 is zero the rod does not press against the wall. This result, which seems at first improbable, may be verified by allowing A to rest on a board on a tank of water and hanging B by a cord; the cord will be found to be vertical when tested by comparison with a plumb line.

105. Special Cases of Equilibrium.—1. *When two forces are in equilibrium they must be equal and opposite and in the same line.* If not equal and opposite they would produce translation, and if not in the same line they would produce rotation.

For example, a body suspended by a cord must rest so that its center of gravity is vertically below the point of support. This supplies an experimental method of finding the center of gravity of a disk of any shape. It is only necessary to support it in succession at two points on its rim and find the intersection of the lines of support.

2. *When three forces are in equilibrium they must all lie in one plane.* For the sum of the moments of all three about any axis is zero. About any axis that intersects the lines of action of two of the forces the moments of these two forces are zero. Hence any such axis must also intersect the line of action of the third force (unless it be parallel to it) and this cannot be so unless all the forces lie in one plane.

3. *Three forces in equilibrium must either be parallel or pass through a single point.* If they are parallel, one is equal and opposite to the resultant of the other two. If they are not parallel, two of them intersect and their moments about the point of intersection are zero. Hence the third must pass through the point of intersection of any two.

As an example of three parallel forces in equilibrium consider (2) of §104. The resultant of F_2 and F_3 must be equal and opposite to and in same line as W which acts at the middle of AB. Hence F_2 and F_3 are equal.

As an example of three non-parallel forces in equilibrium consider (1) of §104. Let the resultant of F_2 and F_3 be F. Then F, F_1 and W are three forces in equilibrium. Hence F must pass through the intersection of F_1 and W. Hence the direction of F is readily found graphically. We may also find graphically the magnitudes of F_1 and F. Since DA and BH are equal, $HBDA$ is a parallelogram. Hence F, F_1 and W are proportional to HA, HB and HD.

106. Stable, Unstable and Neutral Equilibrium.—A body is in equilibrium when it is either at rest or moving uniformly, that is,

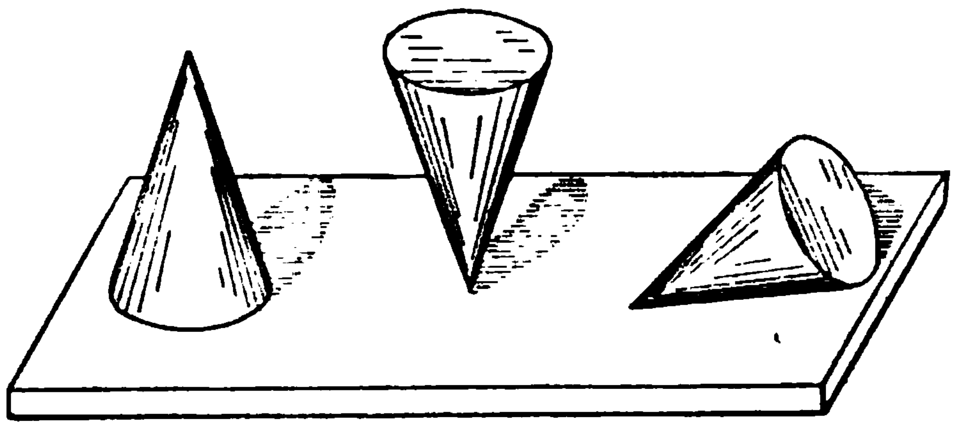


Fig. 53.—Stable, unstable and neutral equilibrium.

without acceleration linear or angular. The resultant of the forces acting on such a body is zero.

When a body in equilibrium is at rest the equilibrium is described as *static*. Of this kind of equilibrium there are three forms, stable, unstable and neutral. A body at rest is in *stable equilibrium* when, on being slightly displaced, it tends to return to its equilibrium position. This is illustrated by a chemical balance, a pendulum or picture hanging by a cord, a book on a table and in fact by most stationary objects. A body at rest is in *unstable equilibrium* when, on being slightly displaced, it tends to move further from its equilibrium position. An egg on end and a board balanced on one corner would be in unstable equilibrium. A body at rest is in *neutral equilibrium* when, on being slightly displaced, it has no tendency either to move further away or

to return; for example, a sphere or cylinder on a horizontal table and any body mounted on an axis through its center of gravity.

A body in a position of stable equilibrium oscillates about that position when displaced and released, though the oscillation may be quickly destroyed by friction or other forces. When too far displaced such a body may come to a position of unstable equilibrium and not return; a table or chair tilted too far comes to a position of unstable equilibrium. The extent to which any such body may be displaced and yet return is a measure of the degree of stability of the equilibrium.

107. Energy Test of Static Equilibrium.—When a body at rest is in stable equilibrium, a disturbance will increase its potential energy. This is evident in the case of a pendulum at rest, for a disturbance raises its center of gravity; work is done against gravity when the body is displaced and this work produces potential energy. Thus *a position of stable equilibrium is a position in which the potential energy is a minimum.* This statement holds true whatever the force against which work is done; the fact that the disturbance produces an increase of potential energy shows that there are conservative forces opposing the motion and these forces will cause the body to return when it is displaced.

A position of unstable equilibrium is a position in which the potential energy is a maximum, as is illustrated by a spheroid on end or a board balanced on a corner; a disturbance lowers the center of gravity. The statement is true whatever the forces in action; for the fact that the body when disturbed moves farther away from its position of equilibrium and thus gains kinetic energy shows that its potential energy diminishes.

When the equilibrium is neutral a displacement produces no change of potential energy; when a sphere rolls on a horizontal table its center neither rises nor falls. An interesting illustration is afforded by the apparatus sketched in the adjoining figure. It will remain at rest whatever the positions of the equal weights which are adjustable along the horizontal rods, for the total potential energy is the same in all.

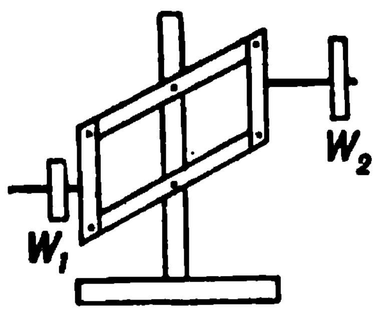


FIG. 54.—Neutral equilibrium.

A body, such as a fly wheel or a railway car, in a steady state of motion is in kinetic equilibrium since the resultant of the forces acting on it is zero.

The principle that for stable equilibrium the potential energy is a minimum is extensively illustrated in nature; the potential energy may be partly or wholly other than mechanical energy, in forms dealt with in other parts of Physics. Changes are continually taking place in nature and bodies, when disturbed, settle into states of stable equilibrium, that is, of minimum potential energy.

KINEMATICS AND DYNAMICS

PERIODIC MOTIONS

108. A periodic motion is one that is repeated in successive equal intervals of time. The time required for each such repetition is called the *period* of the motion. Thus, the moon revolves around the earth with a periodic motion, the period of which is a lunar month and the earth revolves about the sun in a period of a year. The end of a hand of a clock has a periodic motion about the center of the face. A point on a vibrating violin string or piano wire has a periodic motion.

109. Uniform Circular Motion.—When a point P revolves with constant speed in a circle of center O, the position of P at any moment may be assigned by giving the angle that OP makes with some fixed diameter such as OA. This angle is called the *phase* of P's motion.

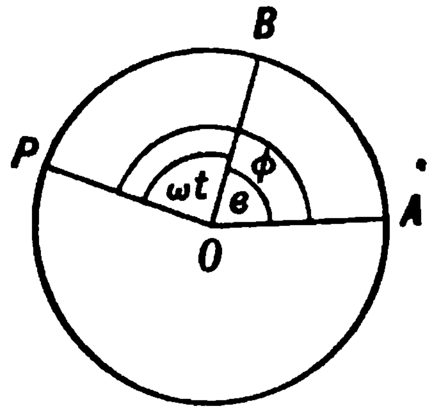


FIG. 55.

If the period of the motion is T, the angle through which OP revolves in unit time is the angular velocity ω and equals $2\pi/T$. Let us suppose that at the moment from which we begin reckoning time P is at some position B, and let its phase at that moment, that is, the angle BOA, be e. After time t, P will have revolved through an angle ωt or $(2\pi/T)t$ and the phase at time t will be $[(2\pi/T)t+e]$.

110. Simple Harmonic Motion.—This is the most important form of periodic motion and is illustrated by the vibration of a simple pendulum swinging in a small arc, of a weight hung from an elastic cord or spring and moving vertically, of a point on the prong of a tuning fork and many other cases.

Simple harmonic motion is a *linear vibration, the motion being such that the vibrating point has an acceleration which is toward the center of its path and proportional to its distance from the center.*

Let $A'A$ be the path of vibration of a point M which has a simple harmonic motion, and let C be the center of $A'A$. Denote the distance of M from C at any time by x, and let values of x be considered as positive when M lies between C and A and negative when M lies between C and A'. When x is positive the acceleration, a, of M is toward C and is, therefore, in the negative direction, and when x is negative a, being still toward C, is positive. Hence, if we denote the constant of proportionality of the magnitude of a to x by c, by the above definition of simple harmonic motion

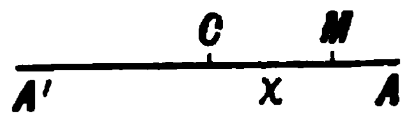


FIG. 56.

$$a = -cx$$

The distance x of the vibrating point from the center of motion is called the **displacement** *of the point.*

One-half of the length of the path of vibration is called the **amplitude** *of the simple harmonic motion.* We shall denote it by r. It is equal to the magnitude of the greatest displacement (CA or CA').

The time required for a complete vibration (that is, from A to A' and back to A) is the **period** of the simple harmonic motion.

111. The Force Acting on a Body that has Simple Harmonic Motion.—A body that has a simple harmonic motion has a varying acceleration which is always directed toward the center. To produce this acceleration a varying force, also directed toward the center, must act on the body. Denote the force by F and let m be the mass of the body. From the Second Law of Motion and the definition of simple harmonic motion we get

$$\begin{aligned} F &= ma \\ &= -mcx \end{aligned}$$

Since m and c are constants for a given body and a given simple harmonic motion, *the force* required is always *opposite to and proportional to the displacement.*

The force required to stretch or compress a spiral spring, one end of which is fixed, is proportional to the displacement of the free end from its unstrained position, and the reaction exerted by the spring is opposite to and proportional to the displacement (§56). Hence a body attached to such a spring and allowed to

vibrate under the action of the spring has simple harmonic motion. The same law of force holds for a flat spring when bent and, in fact, for any elastic body when distorted. Hence all elastic vibrations are simple harmonic motions or compounded of such motions and the same is true of the vibrations that constitute sound and light.

112. Energy of a Body that has Simple Harmonic Motion.—The principle of the Conservation of Energy applies to a body that has simple harmonic motion, since the only force acting on the body is one that depends on the position of the body (§63). We have, in fact, already found in §61 the proper expression for the total energy of such a body in any position; all we need to do is to substitute for k its value in the present case, namely, mc. Hence the total energy is $(\frac{1}{2}mv^2+\frac{1}{2}mcx^2)$, of which the first part is the kinetic energy at displacement x and the second is the potential energy. At one end of the path of vibration v is zero and $x=r$; hence the total energy is potential and equal to $\frac{1}{2}mcr^2$. At the center x is zero and v has its largest value, V; hence the energy is entirely kinetic and equal to $\frac{1}{2}mV^2$.

113. Velocity in Simple Harmonic Motion.—From the result just stated we can find a useful expression for the velocity at any displacement. Equating the total energy at displacement x to that at maximum displacement we have

$$\frac{1}{2}mv^2+\frac{1}{2}mcx^2=\frac{1}{2}mcr^2$$
$$\therefore v=\pm\sqrt{c}\sqrt{r^2-x^2}$$

and, referring to Fig. 56, it will be seen that the positive sign must be taken for motion from A' to A and the negative for motion from A to A'.

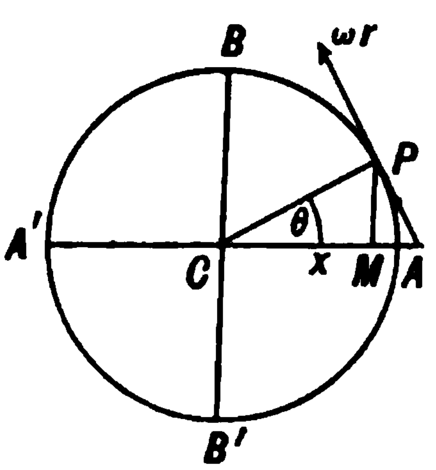


FIG. 57.

114. Relation Between Simple Harmonic and Circular Motions.—Simple harmonic motion has been defined as a vibration in a line according to the law $a=-cx$. Now the projection of a uniform circular motion on a diameter of the circle has exactly the same character. For on $A'A$ as diameter draw a circle, and let P revolve with constant angular velocity, ω, in the circle. If M is the projection of P on $A'A$, M vibrates once along $A'A$ in each revolution of P.

Since the motion of M is that part of the motion of P which is in the direction of $A'A$, the acceleration, a, of M is the component of the acceleration of P in that direction. The accceleration of P is $\omega^2 r$ in the direction PC or $-\omega^2 r$ in the direction CP. Hence

$$a = -\omega^2 r \cos \theta = -\omega^2 x$$

Since ω is constant throughout the motion, the projection is a simple harmonic motion in which $c = \omega^2$. Hence any simple harmonic motion may be regarded as a projection of a uniform circular motion. The circle is called the *circle of reference* of the simple harmonic motion.

This relation between the two kinds of motion affords a means of deducing some of the properties of simple harmonic motion without the use of advanced mathematics.

115. Period of a Simple Harmonic Motion.—Considei a simple harmonic motion as a projection of a uniform circular motion. The periods of the two motions must be the same. Denote it by T. From §114

$$a = -\omega^2 x = -\left(\frac{2\pi}{T}\right)^2 x$$

Hence

$$T = 2\pi\sqrt{-\frac{x}{a}}$$

Since x and a are always of opposite signs, the quantity under the radical is always numerically positive.

116. Trigonometrical Expression for the Displacement.—From the same relation we can also deduce an expression for the displacement at any moment in the simple harmonic motion. Let P be the point whose motion projects into that of the vibrating point M. The angle PCA or the phase of P's motion (§109) equals $[(2\pi/T)t+e]$. Hence for CM or the displacement, x, in the simple harmonic motion, we have

$$x = r \cos\left(\frac{2\pi}{T}t+e\right)$$

While we have deduced this expression from the related circular motion, it must now be regarded as an expression for the simple harmonic motion of amplitude r and period T. $[(2\pi/T)t+e]$ is called the *phase* of the simple harmonic motion at time t, e being the phase of the simple harmonic motion at zero time.

6

For two particular values of e the expression for x becomes simpler. If e is zero, which, as we see from the circular motion, means that at zero time M is at A. the expression for x is

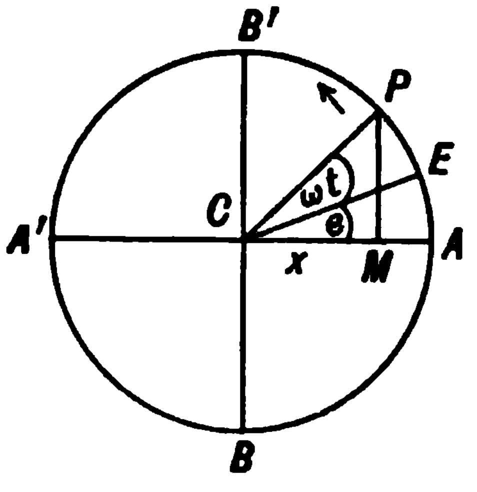


FIG. 58.

$$x=r\cos\frac{2\pi}{T}t$$

If $e=-(\pi/2)$, at zero time P is at B and M is therefore at C and moving in the positive direction. Substituting this value of e in the above general expression for x, we get

$$x=r\sin\frac{2\pi}{T}t$$

117. Simple Pendulum.—A simple pendulum consists of a small heavy body, called the bob (usually spherical), suspended by a practically inextensible cord, the mass of which is so small as to be negligible compared with the bob. As the pendulum swings through a small angle, the bob vibrates through a small arc of a circle which is very nearly a straight line.

The force of gravity, mg, on the bob of the pendulum acts vertically, and it may be resolved into a component along the tangent and a component along the radius. The latter component produces a tension on the cord which does not affect the motion, while the former component produces an acceleration along the tangent. When the cord is at an inclination θ to the vertical, the component along the tangent equals $mg\cos[(\pi/2)-\theta]$ or $mg\sin\theta$. Since the pendulum is supposed to vibrate through a very small angle, $\sin\theta$ may be replaced by θ; in fact, for values of θ less than 2°, $\sin\theta$ and θ are equal within one part in 10,000. If the distance of the bob from its lowest point, measured along the tangent, be denoted by x and the length of the pendulum by l, $\theta=x/l$ radians. Hence the force along the tangent is $mg(x/l)$.

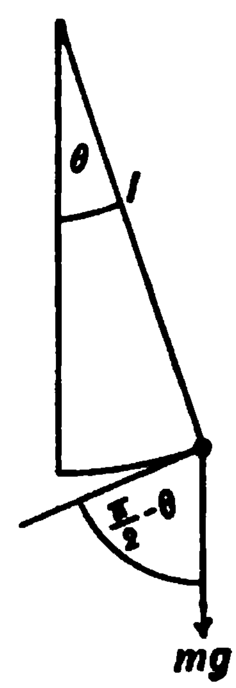


FIG. 59.—Simple pendulum.

This force is in the negative direction when x is positive. Hence, denoting the acceleration along the tangent by a, we have by the Second Law of Motion

$$-mg\frac{x}{l}=ma$$

Hence
$$a = -\frac{g}{l}x$$

Since the multiplier of x is a constant, the acceleration is opposite to and proportional to the displacement. Hence the motion is simple harmonic motion, and, if T be the period or time of vibration of the pendulum, by §115

$$T = 2\pi\sqrt{-\frac{x}{a}}$$
$$= 2\pi\sqrt{\frac{l}{g}}$$

118. Angular Harmonic Motion.—A body attached to an axis may vibrate backward and forward through an angle, as in the case of a balance wheel of a watch or of any heavy body hung on a peg. When the angular acceleration, α, is always opposite to and proportional to the angular displacement, θ, the motion is called *angular harmonic motion*. Hence the general formula for such motion is

$$\alpha = -C \cdot \theta$$

C being a constant.

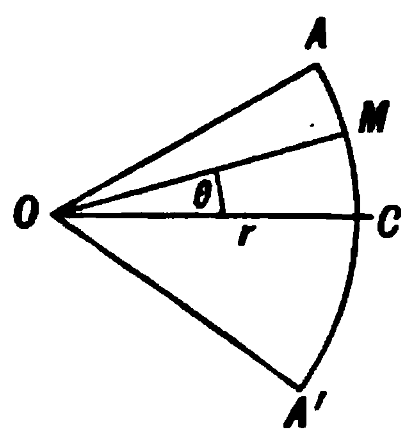


FIG. 60.

Let Fig. 60 be a plane through the body perpendicular to the axis O. A line OM in the body will vibrate backward and forward through an angle. The point M will vibrate in an arc of a circle of radius OM or r. When the angular displacement of OM from its mean position OC is θ, the displacement, x, of M from C is $r\theta$ and the linear acceleration, a, of M is $r\alpha$ (§ 69). Substituting these values of θ and α in the above formula and cancelling r, we get

$$a = -C \cdot x$$

Thus the motion of M is simple harmonic motion in all respects except that it is along an arc (which may be long or short) instead of along a straight line. We might suppose the arc straightened out without any other change in the nature of the motion of M. Hence, if T be the period of M's motion, which, of course, is the same as the period of the angular harmonic motion,

$$T = 2\pi\sqrt{-\frac{x}{a}}$$
$$= 2\pi\sqrt{-\frac{\theta}{\alpha}}$$

This expression for the calculation of the period of an angular harmonic motion is similar to that for the calculation of the period of a simple harmonic motion (§115).

As examples of angular harmonic motion we shall consider the torsion endulum and the physical pendulum.

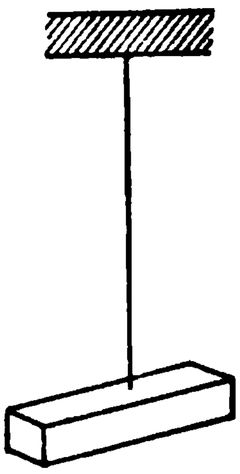


Fig. 61.—Torsional pendulum.

119. The Torsion Pendulum.—A torsion pendulum consists of a vertical wire carrying a body at one end and clamped at the other end. When the body is turned around the wire as axis and released it performs angular vibrations; the twisted wire begins to untwist and thus starts the motion which persists after the wire has untwisted, owing to the kinetic energy acquired by the body.

To twist the wire requires the application of a couple. The twist, θ, produced by a certain couple of moment L, is proportional to L and to the length l of the wire. Hence $Ll=\tau\theta$, where τ is a constant which is called the *constant of torsion* of the wire. The couple exerted by the twisted wire is equal and opposite to that required to produce the twist. Hence the couple exerted by the wire on the body is $-\tau(\theta/l)$ when the displacement is θ. This couple gives the body an angular acceleration, and, if we denote this by α and the moment of inertia of the body by I,

$$-\tau\frac{\theta}{l}=I\alpha$$

$$\therefore\ \alpha=-\frac{\tau}{Il}\theta$$

In this the multiplier of θ is a constant which depends on the wire and the body and is independent of the motion. Hence the motion agrees with the definition of angular harmonic motion, and, if T is the period of vibration,

$$T=2\pi\sqrt{-\frac{\theta}{\alpha}}$$

$$=2\pi\sqrt{\frac{Il}{\tau}}$$

It should be noticed that we have not assumed the angle of vibration to be small, as in the case of the ordinary pendulum; in the torsion pendulum the restoring couple is proportional to

the angular displacement, even when the latter is large (provided it is not so large as to permanently strain the wire).

By means of the torsion pendulum the moment of inertia of an irregular body can be compared with that of a body of known moment of inertia. The two are, by the above formula for T, proportional to the squares of the corresponding times of vibration when the bodies are in turn attached to the same wire and set into angular vibration.

120. The Compound Pendulum.—A body of any shape suspended by a horizontal axis and vibrating under gravity through a small angle constitutes a compound pendulum. Fig. 62 represents a vertical section through the center of gravity C and perpendicular to the axis of suspension S. Denote SC by h. When SC is inclined at an angle θ to the vertical, the force of gravity, mg, which acts at C, has a moment about S equal to $mgh \sin\theta$, which is negative when θ is positive. Hence, if I is the moment of inertia of the body about S,

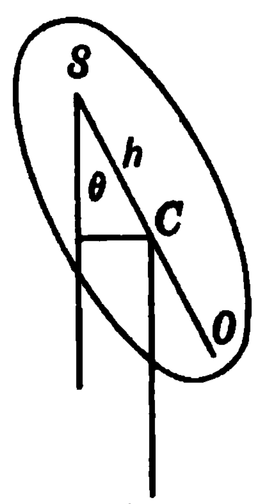


FIG. 62.

$$-mgh \sin\theta = I\alpha$$

If the angle θ is always small, we may, as in the case of the simple pendulum, replace $\sin\theta$ by θ and thus get

$$\alpha = -\frac{mgh}{I}\theta$$

This satisfies the condition for angular harmonic motion and the period of vibration is

$$T = 2\pi\sqrt{-\frac{\theta}{\alpha}}$$
$$= 2\pi\sqrt{\frac{I}{mgh}}$$

Let the radius of gyration about an axis through C parallel to the axis of suspension be k. Then the moment of inertia about the axis through C is mk^2 and about the parallel axis through S it is mk^2+mh^2 (§§83, 85), which is, therefore, the value of I. Hence

$$T = 2\pi\sqrt{\frac{k^2+h^2}{gh}}$$

If this be compared with the formula for a simple pendulum, it is seen that, if l be the length of a simple pendulum that vibrates in the same time as the compound pendulum,

$$l = \frac{k^2+h^2}{h} = \frac{k^2}{h}+h$$

Hence $$(l-h)h=k^2$$

The length l is evidently greater than h. Hence, if we measure along SC a length equal to l, we shall arrive at a point O in SC extended. The point O, which is always on the opposite side of C from S, is the point at which the whole mass of the body might be supposed concentrated without any alteration of the period of vibration. O is called the center of oscillation corresponding to the axis of suspension S. Since $CO=(l-h)$ and $CS=h$, we have as the relation between any center of oscillation and the position of the corresponding axis of suspension

$$CS \cdot CO = k^2$$

If the pendulum be now inverted and set to vibrate about an axis through O parallel to the former axis, the new center of oscillation, O', will lie in OC produced and must satisfy the relation

$$CO \cdot CO' = k^2$$

A comparison of these two equations shows that O' must coincide with S. Hence *the center of suspension and the center of oscillation are interchangeable and the distance between them is the length of the equivalent simple pendulum.* This is the principle of Kater's pendulum.

121. Energy Changes.—The resultant force of gravity acts at C (Fig. 62). Hence the potential energy of the pendulum in any position is the same as if its mass were concentrated at C. But the pendulum does not swing as if it were concentrated at C, because its kinetic energy is that of its mass supposed concentrated at C plus its kinetic energy of rotation about C (§91). As the pendulum falls toward the vertical the lost potential energy goes partly into energy of rotation about C; hence it does not swing as rapidly as if it were concentrated at C, that is, as if it were a simple pendulum of length SC. A parallel case that brings out the distinction is illustrated by a block suspended by two cords as in Fig. 63. Swinging perpendicularly to the plane of the figure it is a physical pendulum of length SO, the block having energy of rotation. Swinging parallel to the plane of the figure it is a simple pendulum of length equal to the length of the cords; the block in this case has no rotation. A similar explanation applies to the motion of the pans of a balance. They do not rotate with the beam but move vertically; hence they affect the motion as if concentrated on the supporting knife-edges.

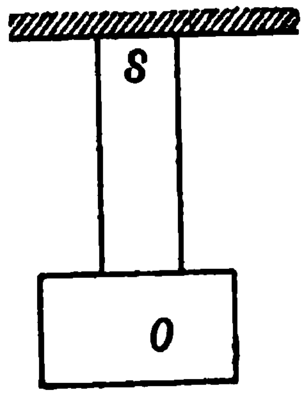


FIG. 63.

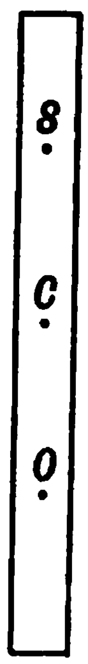


FIG. 64.—Center of percussion.

122. Center of Percussion. There is another important relation between an axis of suspension S and the corresponding center of oscillation O. A blow at O transverse to SO will start the body rotating about S without any jar on the support at S. Hence O is also called the *center of percussion* of the body when suspended at S. The center of percussion is readily found by suspending the body by a cord

and striking horizontal blows at various points. Or it may be found by holding the body at S and striking across a table edge, as a base-ball player strikes a ball with a bat; when the blow is through the center of percussion there is no jar on the hand.

123. Gyroscopic Motion.—A gyroscope is a wheel on a horizontal axle which is supported on a pivot (a bicycle wheel suspended by a vertical cord attached to a short extension of the axle will serve). When the wheel is set in rotation and the axle then released, the axle, instead of tilting in a vertical plane, as it would if the wheel were at rest, revolves in a horizontal plane at a rate that depends on the velocity of rotation of the wheel about the axle. This motion is called **precession.** (Slight vertical oscillations or *nutations* of the free end of the axle may also accompany the precession.) The weight of the

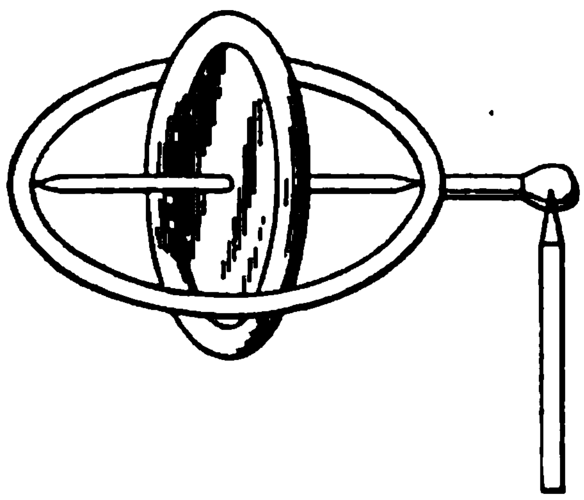


FIG. 65.—A gyroscope.

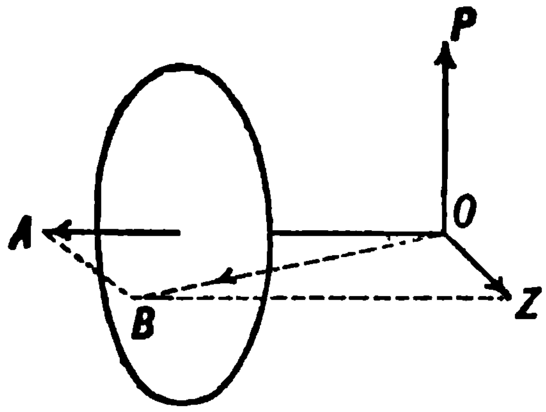


FIG. 66.

wheel acting at the center of the wheel has a moment about an axis through the pivot at right angles to the axis of the wheel. If this moment of force be increased by hanging a weight on the frame the rate of precession will be greater. If the wheel be supported at its center of gravity there will be no moment of force and no precession. (Thus mounted the instrument is sometimes called a gyrostat.)

If the motion be carefully considered it will be seen that it is very analogous to the revolution of a particle in a circle under the action of a force directed toward the center. The latter requires a force *perpendicular to the direction of motion,* while precession requires a moment of force about an axis *perpendicular to the axis of rotation.*

124. Moment of Force Required to Produce Precession.—To find the magnitude, L, of the moment of force required for precession, let OA

represent the angular momentum, $I\omega$. After a short time, t, OA will have turned through a small angle ϕ to the position OB, where $\phi = \omega' t$, ω' being the angular velocity of precession. Hence angular momentum represented by OZ or AB must have been added and this must equal Lt (§87).

$$\therefore \frac{Lt}{I\omega} = \frac{AB}{OA} = \omega' t.$$

Hence $$L = I\omega\omega'$$

Whenever a rotating body shows precession, that is, when its axis of rotation is revolving, some agent must be applying the moment of force, L, about an axis perpendicular to those of rotation and precession and must be experiencing an equal and opposite reaction.

125. Other Examples of Precession.—The curvature of the path of a coin rolled with a tilt along a table is due to the precession of its axis caused by the moment of its weight about the point of contact with the table. The motion of a top is a precession.

Any large body, such as a dynamo armature, in rotation aboard a vessel that is rolling, pitching, or turning, has a precessional motion and the bearings must supply the necessary moment of force and experience an equal and opposite reaction.

When a side-wheel steamer is turned in a sharp curve there is a precession of the axis of the paddle wheels. To produce this precession and at the same time keep the vessel level would require a moment of force about a longitudinal axis, and in the absence of such a moment the vessel lists to the outer side.

The earth is not quite spherical but bulges at the equator. On one side the protuberance is closer to the moon than the center of the earth and on the other side it is farther away. The result of this (and of a similar but smaller moment exerted by the sun) is a moment of force that causes a precession of the earth's axis.

The gyroscope has been applied to steering torpedoes, to preventing the rolling of ships, to balancing trains on a single rail, and in the construction of a non-magnetic mariner's compass.

FRICTION

126. Static Friction.—When two solids are in contact there is a resistance, caused by the surfaces, to the sliding of one on the other. This resistance is called *Friction*. When a force parallel to the surfaces of contact is applied to one of the bodies and the

force is less than a certain amount, which depends on the nature of the surfaces and the pressure between them, motion will not take place, the resistance being equal to the force. When the force is increased to a certain value the resistance will fail to increase and sliding will take place. This maximum resistance is called the *maximum static friction.* With a given pair of surfaces in contact and with a force tending to produce sliding motion in a certain direction (to take account of the influence of grain), *the maximum static friction is found to be* (within certain wide limits of pressure) *proportional to the pressure.* Denoting the coefficient of proportionality by μ, the maximum of static friction by F, and the pressure by P, we have

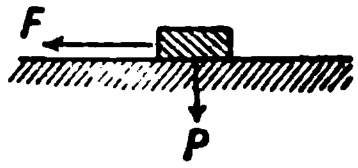


FIG. 67.

$$F=\mu P$$

The constant μ is called the *coefficient of static friction*, and may be defined as *the ratio of the maximum static friction between two surfaces to the pressure between them.*

By the pressure here is meant the total perpendicular force between the surfaces (not the force per unit of area, as the word pressure is sometimes used). If one of the two bodies rests on the other and if the surfaces of contact are plane and horizontal, the pressure is the weight of the upper. The maximum static friction is the force applied horizontally to the upper that will just produce motion. If additional weights be placed on the upper body, the pressure between the surfaces will be increased and the friction will be increased in the same proportion. If the upper mass be redistributed in any way, for instance, if it be cut in two and one part placed on the other, the total force of friction will not change; for, while the area of contact will be diminished, the pressure on each unit of area will be increased in the same proportion. Thus *the total frictional resistance is independent of the area of contact* and for two given surfaces depends only on the pressure, as is implied in the equation $F=\mu P$.

127. The coefficient of static friction between two surfaces depends on the materials and a variety of circumstances. The rougher the surfaces, that is the greater the inequalities in each, the larger is μ. If the surfaces are not clean parts of the surfaces are replaced by surfaces of the foreign

substance and μ is necessarily different. The longer two surfaces are in contact the greater the maximum static friction; this is especially true of soft or fibrous surfaces. When the materials are of grained structure the friction is greater across the grain than along it. Friction is, no doubt, due to interlocking of the projections on one surface with those on the other surface. When slipping takes place some projecting pieces are broken off or abraded as it is called. With prolonged contact between two surfaces small readjustments of the surface particles take place, so that the fit becomes closer and the resistance to motion greater. It has even been found that when one surface is pushed a *very* small distance it will when released spring back, thus showing that there is some elastic bending of surf aceprojections. In general, friction between two surfaces of the same material is greater than between surfaces of different material since the former allows more uniform interlocking. Thus there is an advantage in using brass bearings for steel shafts to diminish friction, and covering with leather the face of a pulley used with leather belting increases friction and helps to prevent slip.

Friction is utilized in the transmission of energy by machine belting. Usually some slipping takes place, for the belt stretches somewhat while in contact with the pulley. Friction between the driving wheels of a locomotive and the rails prevent slipping; without it the locomotive would be helpless, and where it is not sufficient the track is sanded. To hold a rope fast it is sometimes wrapped around a post. The friction on each part of the rope diminishes the tension transmitted to the next part. It is found that after one turn the tension is diminished to about $\frac{1}{6}$, after two turns to $\frac{1}{6}$ of $\frac{1}{6}$ and so on. At this rate after five turns a pull of one pound weight on the free end would counteract a force of 4 tons at the other end.

The laws of friction were first investigated by Coulomb and are sometimes called by his name.

128. Slip on an Incline.—When a body rests on an inclined plane the tilt of which is gradually increased there is some angle i at which slipping begins.

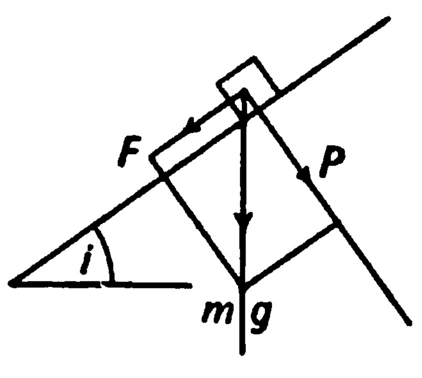


Fig. 68.

The weight of the body is mg and acts vertically. It may be resolved into a component $mg \sin i$ down the plane and a component $mg \cos i$ perpendicular to the plane. The latter component causes pressure between the surfaces, while the former is the force parallel to the surface which produces motion. Hence from the definition of this coefficient of static friction

$$\mu = \frac{F}{P} = \frac{mg \sin i}{mg \cos i} = \tan i$$

Thus the coefficient of static friction is equal to the tangent of the angle of slip. (This angle is also sometimes called the angle of repose.) This relation provides a simple method of measuring μ.

129. Kinetic Friction.—To keep one body sliding on another at a constant speed a certain force, F, parallel to the surface of contact is required. Through a considerable range of speed this force is practically constant. The opposing resistance offered by the surfaces is called *kinetic friction*. It is found to be, for a certain pair of surfaces moving in a definite direction, proportional to the pressure, P, between the surfaces. Denoting the coefficient of proportionality by μ', we have

$$F = \mu' P$$

The constant μ' is called the *coefficient of kinetic friction*. It may be defined as *the ratio of the kinetic friction between two surfaces to the pressure between them.*

As in the case of static friction, for a given pressure between two surfaces the kinetic friction is independent of the area of contact.

The kinetic friction between two surfaces is in general less than the maximum static friction. The reason probably is that time is not allowed for the surface to settle into as close contact as if they were at rest. Moreover, kinetic friction is not quite independent of velocity. When the velocity is decreased until it is very small (how small depends upon the particular surfaces) the friction increases and it continues to increase as the velocity diminishes toward zero, and at a sufficiently small velocity the kinetic friction probably does not differ appreciably from the maximum static friction. At very great velocities the friction is generally less than at moderate velocities.

A friction dynamometer is a machine for measuring the power of an engine; the engine drives a wheel over which a belt hangs under known tension. From the tension of the belt and the number of revolutions made by the latter the work done is calculated.

When a lubricant is used between two surfaces there is no longer friction of solid on solid and the laws of kinetic friction no longer hold; the coefficient of friction depends on both pressure and velocity and the action is very complex. The friction of a skate on ice is probably greatly diminished by the momentary liquefaction of the ice immediately under the

skate due to the great pressure exerted by the latter on a small area (see §301").

130. Sliding on an Inclined Plane.—A body sliding down an inclined plane (Fig. 21) is urged downward by the component of its weight along the plane and retarded by friction. If the inclination of the plane to the horizontal is i, the component of gravity along the plane is $mg \sin i$. The pressure perpendicular to the plane is $mg \cos i$; hence the force of friction is $\mu' mg \cos i$. If the component of gravity down the plane exceeds the force of friction, the body will slide with an acceleration a. Hence, taking the direction down along the plane as the positive direction, we have by Newton's Second Law

$$ma = mg \sin i - \mu' mg \cos i$$

This suggests a method of finding μ' by measuring a and i.

131. Rolling Friction.—The term friction is also applied to the resistance experienced by a wheel in rolling on a surface without any slipping. The cause of the resistance is in this case entirely different. This is seen by considering the rolling of a heavy wheel on a soft substance, such as India rubber. If the wheel were at rest it would sink into the rubber, raising a small mound on each side of the contact. When the wheel is moving forward the mound is chiefly on the forward side at A. The pressure, P, of the rubber on the wheel at A is inclined to the vertical, in some such direction as AP. The point about which the wheel is momentarily rotating (§69) is not C but B in the figure, and the moment of P about B is necessarily opposed to that of F.

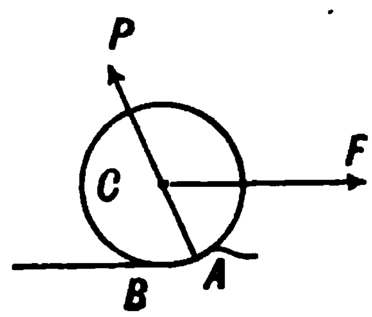


FIG 69.—Resistance to rolling.

It will be noted by the explanation that the resistance to the motion is greater the softer the surface, greater the greater the pressure of the wheel on the surface, and less the larger the wheel, since a larger wheel will distribute its pressure over a larger surface and will not sink so deeply. When the surface on which the wheel rolls is hard very little deformation will ensue, and the resistance to the motion will be much less. Thus the resistance to the rolling of iron on india rubber is about ten times greater than the rolling of iron on iron. With a lignum vitæ cylinder of 16-in. diameter loaded with 1000 lbs. the rolling friction has been found to be

about 3 per cent. of the sliding friction when the wheel was not allowed to rotate. Because of this difference rolling is, when possible, preferred to sliding. Thus rollers beneath a heavy body and the balls in a ball-bearing greatly diminish frictional resistance.

A pneumatic tire on a bicycle or automobile flattens out in contact with the ground and does not sink in, so that it gives the wheel the advantage of a much larger wheel. But it also bulges a little in front of the flattened part and this bulge is an obstruction of the same nature as the little mound in Fig. 69. On a perfectly smooth, plane, hard road a pneumatic tire would be a disadvantage. On a soft rough road it is a great advantage. For a hard smooth road the tire should be pumped "hard" for a soft road it should be "soft."

SIMPLE MACHINES

132. Machines.—A machine is a contrivance for applying energy to do work in the way most suitable for a certain purpose. The machine does not create energy; no machine can do that. To do work it must receive energy from some store of energy, and the greatest amount of useful work it can do cannot exceed the energy it receives.

Different machines receive energy in different forms, some in the form of mechanical (kinetic and potential) energy, some in the form of heat energy, some in the form of chemical energy, and so on. We shall only consider here machines which employ mechanical energy and do work against mechanical forces.

In certain very elementary machines, the so-called *simple machines*, the agent which supplies the energy exerts but a single force and the machine, at least as regards the useful work which it performs, is opposed by a single resisting force. The former is frequently called the "power"; but, to avoid confusion, we shall call it the *applied force*. The resisting force is frequently called the "weight"; but, as the opposing force is not always that of gravity, we shall call it the *resistance*.

Every machine in its action encounters a certain amount of frictional resistance; the work done against it is not usually useful work. This in many cases is very small, and, in treating (to a first approximation) of the simple machines, it is customary to neglect it.

133. Mechanical Advantage.—The work done by the applied force, P, is measured by the product of P and the distance, p,

through which P acts. The work done against the resistance is measured by the product of the resistance, Q, and the distance q, through which it is overcome. In a simple machine (where friction may be neglected) these must be equal. Hence

$$\frac{Q}{P}=\frac{p}{q}$$

Hence p is greater than q in the proportion in which Q is greater than P. This principle was first stated by Stevinus (1548–1620). It is frequently put in the form "*what is gained in power* (*i.e.* *force*) *is lost in speed.*"

The ratio of Q to P for a machine is called the *mechanical advantage* of the machine. Since, for a perfect machine, that is, one in which friction is negligible, the above ratio is also the ratio of p to q, it follows that we can deduce the mechanical advantage of such a machine from the ratio of the speeds without considering the inner mechanism of the machine.

134. Efficiency.—By the *efficiency* of a machine is meant the ratio of the useful work, or work of the kind desired, to the energy received. For a simple machine without friction this would be unity. When there is friction the efficiency may have any value less than unity.

135. Levers.—A lever is a bar supported at a point called the *fulcrum*, F; a force, P, applied to the bar at a point A will over-

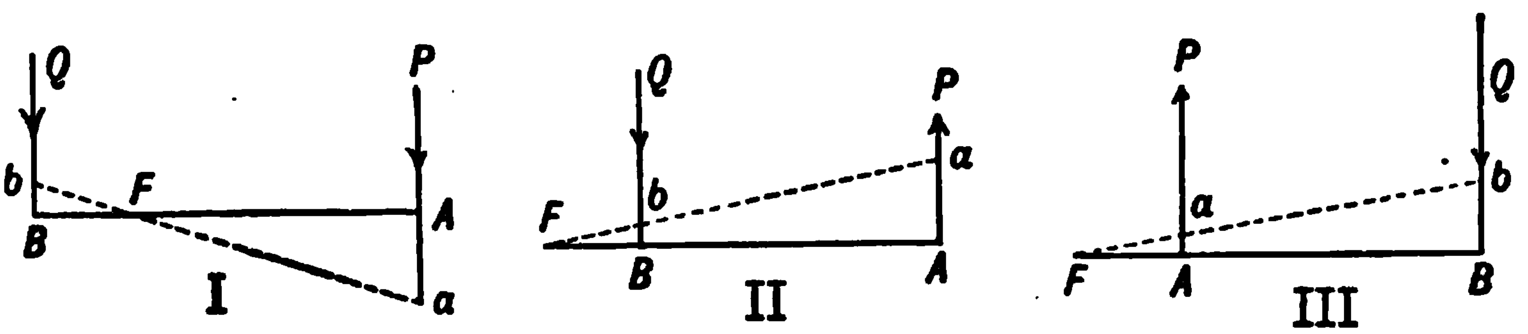


Fig. 70.—Three classes of levers.

come a resistance, Q, acting at another point B. We shall suppose that P and Q act at right angles to the bar and to the axis of rotation at the fulcrum.

To find the relation between P and Q suppose the bar to turn through a very small angle, so that A moves through a distance Aa and B through a distance Bb. The work done by P is $P{\cdot}Aa$

and the work done against Q is $Q \cdot Bb$. The conservation of energy requires that these should be equal. Hence

$$\frac{Q}{P}=\frac{Aa}{Bb}=\frac{AF}{BF}$$

This relation may also be found by considering the parallel forces acting on the bar or by taking moments about the fulcrum.

Levers are usually divided into three classes represented by the figures. In levers of the *first class* the force, P, and the resistance, Q, are on opposite sides of the fulcrum, and the resistance may be greater or less than the applied force. To this class belong the crow-bar, forceps, scissors, poker, and the common balance.

In levers of the *second class* the applied force and the resistance are on the same side of the fulcrum, the former being farther from it than the latter. Thus the resistance is always greater than the applied force. This class includes the oar of a boat, a pair of nut-crackers, a claw-hammer for extracting nails, etc.

In levers of the *third class* the applied force and the resistance are on the same side of the fulcrum, the former being nearer to the fulcrum than the latter. The purpose of such a lever is a gain of displacement or of speed. This class includes the forearm which is hinged at the elbow and acted on by the biceps at a distance of two or three inches from the elbow, a pair of tongs and the lever of a safety-valve for steam pressure.

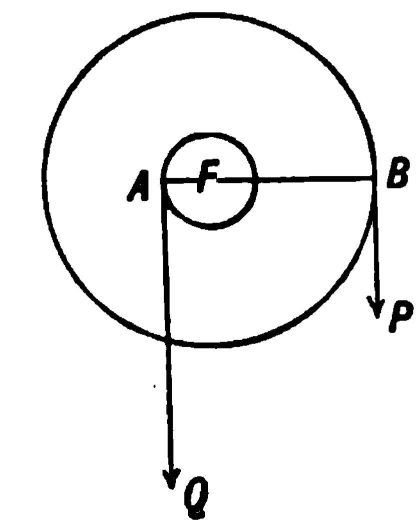


FIG 71.—Wheel and axle.

136. The Wheel and Axle.—A straight lever cannot raise a weight higher above the fulcrum than the distance of the weight from the fulcrum. The apparatus called a "wheel and axle" acts on the same principle as a lever but its range is not so limited. It consists of a wheel of large radius rigidly connected to an axle of smaller radius. The applied force, P, acts on a cord wrapped around the wheel, while the weight or resistance acts on a cord wrapped around the axle. The principle involved is that of a lever of the first class, the radius, R, of the wheel being the lever arm for the applied

force, while the radius, r, of the axle is the lever arm of the resistance. Hence

$$\frac{Q}{P}=\frac{R}{r}$$

This formula may also be proved directly by equating the work done by P in one complete revolution, $2\pi RP$, to the work done against Q, $2\pi rQ$; also by taking moments about F.

The principle of the Wheel and Axle is applied in the pilot wheel and in the capstan where the wheel is replaced by spokes in the axle, and in the winch, where there is but a single spoke, the crank arm.

In the above we have neglected friction, which is always considerable.

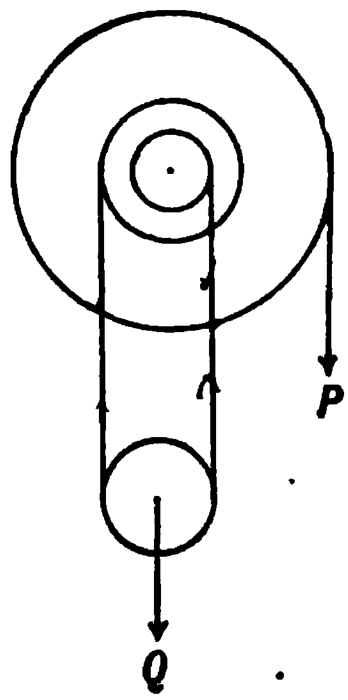


FIG. 72.—Differential wheel and axle.

137. Differential Wheel and Axle.—To obtain a very high mechanical advantage the wheel would have to be made very large, which would be inconvenient, or the axle would have to be made very small, which would greatly weaken it. To avoid these disadvantages the axle is made in two parts of different size and the cord is wrapped in the same direction around both, as indicated in the figure, the weight being carried by a pulley through which the cord passes.

Let the radius of the wheel be R, that of the large part of the axle r, and of the small part r'. The upward force on the pulley is twice the tension of the cord and the downward force is Q, the weight of the pulley being neglected. Hence, by the principle of forces in equilibrium, the tension in the cord is $\frac{1}{2}Q$. In one revolution P does work $P\cdot 2\pi R$ and the tension of the cord acting on the smaller part of the axle does work $\frac{1}{2}Q\cdot 2\pi r'$, while work $\frac{1}{2}Q\cdot 2\pi r$ is done against the tension in the cord acting on the larger part of the axle. Hence

$$P\cdot 2\pi R+\tfrac{1}{2}Q2\pi r'=\tfrac{1}{2}Q\cdot 2\pi r$$
$$\therefore \frac{Q}{P}=\frac{2R}{r-r'}$$

138. Pulleys.—The simplest pulley is a wheel for the purpose of changing the direction in which a force is applied. It consists of a wheel in a framework or block which is either fixed

or free. If it is fixed, the direction of the force is changed without any change in the magnitude (see Fig. 73a).

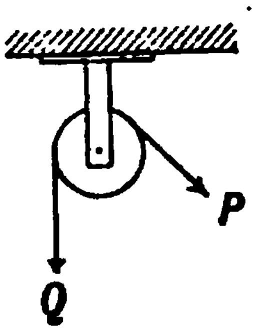


FIG. 73a.

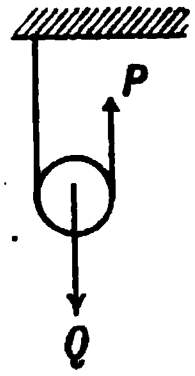


FIG. 73b.

If it is free and the two parts of the cord are parallel, the tension in any part of the cord is (neglecting friction and the weight of the cord) equal to the force applied at its free end. Hence for equilibrium

$$Q=2P$$

If the weight of the pulley is not negligible it may be included in Q. This formula is also readily found by the principle of energy; for each unit of the length that Q moves P must move two.

139. Block and Tackle.—Several pulleys are frequently used in combination so as to secure higher mechanical advantage. The most common arrangement is called the *block and tackle*. The pulleys are in two blocks with several pulleys in each block. The fixed end of the cord may be attached to either the upper or the lower block; if to the former, there will be an equal number of pulleys in the two blocks, as in the figure; if to the latter, there will be one more pulley in the upper block. When the distance between the blocks is decreased by one unit of length, each branch of the cord in contact with the lower pulley must shorten one unit of length. Hence

FIG. 74.—Block and tackle.

$$Q=nP$$

where n is the number of branches of the cord at the lower block.

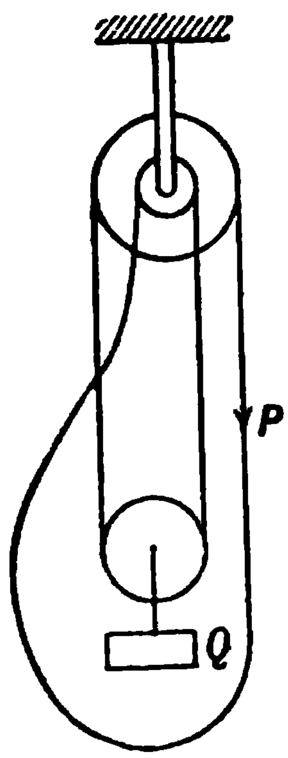


FIG. 75.

140. The Differential Pulley or Chain Hoist.—In this the upper block holds two pulleys of different diameters fixed rigidly to the same axis, while the lower block is replaced by a single pulley. An endless chain passes over the three pulleys as shown in the figure and is prevented from slipping by teeth on the pulleys. This is essentially a modification of the differential wheel and

axle in which the wheel and the larger part of the axle have the same radius. The relation between P and Q, which may be worked out independently or may be obtained by putting $R=r$ in the formula of §137, is

$$\frac{Q}{P}=\frac{2r}{r-r'}$$

141. The Inclined Plane.—A force less than the weight of a body may suffice to draw the body up an inclined plane. Let P be the force and W the weight (Fig. 76a). Also let h be the

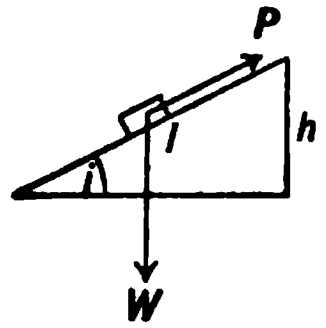


FIG. 76a.

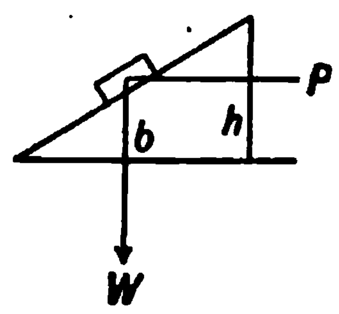


FIG. 76b.

height and l the length of the plane. When the body has been drawn up the whole length of the plane the work done by P (neglecting friction) will be Pl and the work done against W or the increase of potential energy will be Wh. These must be equal. Hence

$$\frac{W}{P}=\frac{l}{h}$$

This is essentially the same expression as already found (§50) by considering the component of W down the plane. If friction cannot be neglected the work done against it will be Fl, where F is the force of friction, and in the above equation P must be replaced by $(P-F)$.

If P act horizontally (Fig. 76b) the work done by P will be Pb. Hence (neglecting friction)

$$\frac{W}{P}=\frac{b}{h}$$

142. The Screw.—The thread of an ordinary screw makes a constant angle with the length of the screw. If the thread of a vertical screw were supposed unwrapped, with its inclination kept constant, it would be an inclined line. The pitch of a screw

is the distance, parallel to the length of the screw, between consecutive turns of the thread. The pitch divided by the outer circumference is the tangent of the inclination of the thread to the length of the screw.

If a nut carrying a heavy weight be turned around a vertical screw so that it ascends, the process will be similar to forcing a heavy body up an inclined plane by a horizontal force. In the jackscrew for raising heavy bodies the nut is fixed while the screw is turned by a lever. The useful work performed by the screw in one turn is the product of the resistance it overcomes, Q, and the rise in one turn, which is the pitch h. The work done in the same time is the product of the applied force, P, and the circumference, $2\pi R$, of the end of the lever arm. Equating these would give us a relation between P and Q; but friction is in general so large as to render the relation inapplicable.

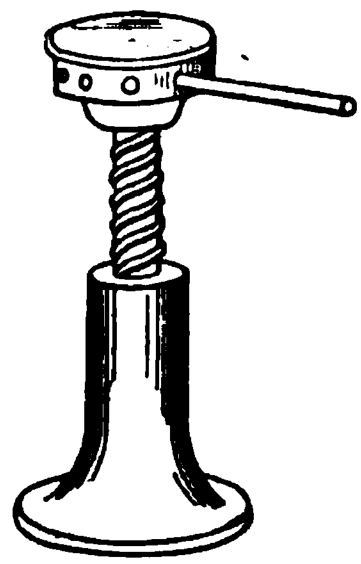


FIG. 77.—Jackscrew.

GRAVITATION

143. Law of Universal Gravitation.—Until the time of Newton the weight of a body, or the measure of its tendency to fall to the earth, was generally regarded as an inherent property of matter that needed no further explanation. To Newton (and to some of his contemporaries) it occurred that the weight of a body on the surface of the earth is due to a force of attraction between the body and the earth, and that this attraction is only a particular case of a universal attraction between all bodies no matter where situated. Newton then sought to discover the law that such a force would have to follow to account for the facts, how it would have to depend on the masses of the bodies and their distance apart. Now it was not possible for him to change the distance between a body and the center of the earth by any except an exceedingly small fraction, and the force between two bodies of ordinary size on the surface of the earth was so small that it escaped detection until a much later date. Hence he turned his attention to the motion of the moon and the planets.

Before the time of Newton, Kepler had, by a very extensive and painstaking study of the motions of the planets, arrived at certain laws known as Kepler's Laws. These may be stated as follows:

1. The areas swept over by a line joining a planet to the sun are proportional to the times.

2. Each planet moves in an ellipse in one focus of which the sun is situated.

3. The squares of the periods of revolution of the planets are proportional to the cubes of the major axes of the ellipses.

From these laws Newton showed that the motions of the planets could be accounted for on the supposition that between each planet and the sun there is a force of attraction, proportional to the product of the masses and inversely as the squares of their distances apart.

Newton also showed that, if we suppose that there is a force according to this law between every two particles, a sphere that is either homogeneous or may be regarded as made up of shells each of which is homogeneous will attract an outside body as if the sphere were concentrated at its center. The earth is very nearly such a sphere and must, therefore, according to the law of gravitation, attract (approximately) as if concentrated at its center.

144. Motion of the Moon.—As evidence for the law of gravitation Newton showed that it correctly accounts for the motion of the moon. At the surface of the earth a body is attracted by the earth as if the latter were concentrated at its center. Now the radius of the earth is approximately 4,000 miles and the average value of the acceleration of a falling body may be taken as 32.2 feet per sec^2. The distance of the moon from the earth, which is somewhat variable, may be taken as approximately 240,000 miles or 60 times the radius of the earth. Hence, according to the law of gravitation, the acceleration of a body at the distance of the moon due to the earth's attraction should be $32.2/60^2$ or .00894 ft. per sec^2.

The acceleration a of the moon towards the earth (§32) equals v^2/R. The period of rotation of the moon, also slightly variable, is about 27 days, 8 hours. Calling this T, we have $v = (2\pi R/T)$. Hence $a = (4\pi^2 R)/T^2$, or reducing R to feet and T to seconds

$a = .00896$. This value of a, calculated from the observed period of the moon, agrees as closely with the preceding value, deduced from the law of gravitation, as could be expected when the fact is considered that only approximate values for the various constants have been used. The argument must be considered very strong evidence for the law of gravitation.

145. Force of Gravitation Proportional to Mass.—According to the law of gravitation the attraction between two bodies is proportional to their masses and is independent of the materials of which they consist. One proof of this was given by Galileo, when he dropped two cannon balls of different sizes from the leaning tower of Pisa and found that they reached the ground in very nearly the same time. Their accelerations being equal, the ratio of the force to the mass, must, according to the second law of motion, be the same for both. Yet in Galileo's experiment the larger weight was slightly ahead of the smaller, and Galileo correctly explained this difference by remarking that the air-friction would be proportionately less on the larger body. In fact, because of this air friction and the rapidity of the motion, it would be difficult to give a very convincing proof of the law by means of bodies falling with the full acceleration due to gravity.

To avoid this difficulty Newton experimented with a pendulum, the motion of which depends on gravity but on a fraction only of the full force of gravity, namely, the component along the arc of vibration. The bob of the pendulum was a thin shell and into this he put in successive experiments different substances. In each case the same *weight*, as tested by weighing with a balance, was put into the box and, since the force of air-friction on the box for the same amplitude of vibration would be the same no matter what the contents of the box, it followed that at a given inclination to the vertical the force causing the motion would be always the same. He found that the time of vibration was always the same no matter what the contents of the box and hence the masses must also have been the same; that is, equal masses of different substances have equal weights. These experiments were afterward repeated by Bessel with much greater care and with the same result.

The above experiments prove that gravitation is not, like magnetic attraction, a force that depends on some quality of a body

other than its mass; that is, not a selective force but a general force. That it does not depend on any other physical condition such as temperature, or on any chemical condition such as molecular combination, has also been shown by most careful weighing. A third body placed between two bodies has not the least effect in shielding them from their mutual attraction. The fact that a lump of gold, when hammered out into an exceedingly thin sheet, suffers no change of weight shows that the weight of a body does not depend on its form, that gravity acts on the particles whether surrounded by other particles of the same kind or not.

146. The Constant of Gravitation.—The law of gravitation may be stated as a formula, viz.

$$F = G\frac{mm'}{r^2}$$

where G is a constant number called *the constant of gravitation.* To find the magnitude of G it is necessary to measure F in some case where m, m' and r are all known. This was first done by Henry Cavendish in 1797-8, and the experiment, usually called the *Cavendish experiment*, has been repeated many times since with increasing care and accuracy. Cavendish suspended two balls, A and B, from the ends of a long light horizontal rod which was supported by a long fine vertical wire attached to the middle, C, of the rod. On opposite sides, horizontally, of the balls and at known equal distances he placed two large spheres of lead, P and Q. The attraction between each ball and the adjacent large sphere had a moment about C that produced a twist of the supporting wire. When the spheres were in the position P_1Q_1, the twist was in one direction, and when they were in the position P_2Q_2 the twist was in the opposite direction. To deduce the force of attraction from the magnitude of the twist, the constant of torsion of the wire (§119) had to be found by timing vibrations of the wire, when the spheres were removed to positions where they had no influence on the

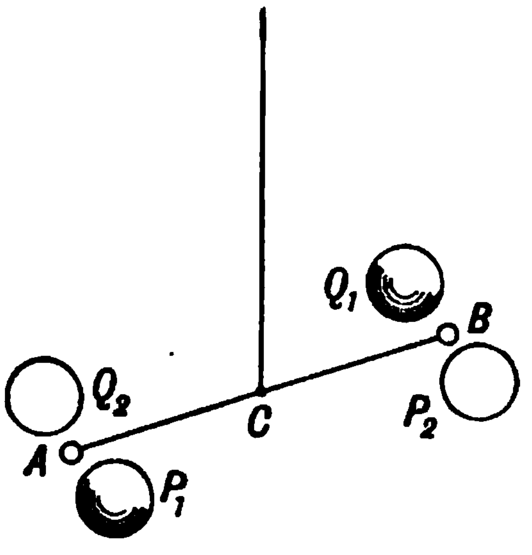


FIG. 78.—Principle of the Cavendish experiment.

vibrations of AB. Thus F, m, m', and r were found and when they were substituted in the above formula the value of G was obtained.

In more recent work the apparatus has been greatly improved. The greatest improvement has been in the substitution of very fine quartz thread for the wire. This also permitted of the apparatus being greatly reduced in size, so that, whereas AB in Cavendish's experiment was 6 ft. long, in Boys' apparatus it was only 0.9 inch, and the masses A, B, and P, Q, were also greatly reduced in size. The value obtained for G (using c.g.s. units) was 6.6579×10^{-8}; this is, therefore, the force in dynes of the attraction between spheres of one gram each at a distance of 1 cm. between their centers. When it is remembered that a dyne is about the weight of a milligram, it is seen that the force measured in the above experiments must be exceedingly small; hence the difficulty of the experiment.

147. The Mean Density of the Earth.—The determination of G made it possible to calculate the mass of the earth (hence Cavendish is sometimes said to have been the first to "weigh the earth"). For if m' in the formula for the law of gravitation be put equal to one gram and m and r be taken as respectively the mass and radius of the earth, F will be the force of attraction between the earth and a body of 1 gm. mass and this is, as we know, 980 dynes. Thus the formula gives us the value of m, the mass of the earth, which is found to be 5.97×10^{27} gms. This figure is so large as to convey no distinct meaning, but a different way of stating the result will be more easily comprehended. The density of a homogeneous body such as water is its mass per unit volume, and when the density of a body is not everywhere the same we may speak of its *mean density* or its whole mass divided by its whole volume. Thus to get the mean density of the earth we divide its whole mass, as given above, by its whole volume. The result is 5.527, that is to say, on the average the earth is 5.527 times as dense as water. It is remarkable that Newton, reasoning from the very slight evidence available in his time, supposed the mean density of the earth to be between 5 and 6.

148. The Tides.—At any place on the shore of an ocean the level of the water rises to a maximum and falls to a minimum once in about every twelve hours and 25 minutes. These risings and fallings are called the *tides*. They are due to the forces of attraction which the moon and the sun exercise on the water on

the surface of the earth and to the rotation of the earth. The complete explanation of their action is extremely difficult, owing to the irregularities of the continents and to other causes.

UNITS

149. Fundamental and Derived Units.—The measurement of any quantity consists in comparing it with a unit of the same kind (§2). Thus a length is measured by comparing it with a unit of length, such as the foot or meter; a velocity is measured by comparing it with a unit of velocity, such as a foot per second and so on. Hence we need as many units as there are different kinds of quantities to be measured.

But all these necessary units are not necessarily independent. It is found that in Mechanics three independent or *fundamental* units are sufficient; all others can be defined in terms of these. A unit defined by reference to some other unit or units is called a *derived unit.*

150. Absolute Systems of Units.—A system of units in which the derived units bear the simplest possible relation to the fundamental units is called an *absolute* system. In such a system the unit of area or surface is the square of the unit of length, the unit of volume is the cube of the unit of length, the unit of velocity is a velocity of unit length per unit time, and so on. Given any three fundamental units of length, time and mass, we can build up an absolute system of derived units. Thus we have one absolute system founded on the cm., gm., and sec., another founded on the ft., lb., and sec., and so on.

151. Dimensions of Units.—It is sometimes necessary to translate results from one absolute system to another. It then becomes necessary to consider how the magnitude of a derived unit changes when the fundamental units are changed. For this purpose we need to know the dimensions of the derived unit, that is, the powers of the fundamental units to which the derived unit is proportional. For instance, the unit of area is the square of the unit of length, or area is of 2 dimensions in length, a statement briefly summarised by the *dimensional formula* $[A]=[L]^2$; similarly, using [Vol] for the unit of volume, $[\mathrm{Vol}]=[L]^3$.

152. Dimensions of Velocity.—The unit of velocity is defined in terms of the unit of length and the unit of time. To find the dimensions in these units consider any relation between velocity, length, and time, such as $s=vt$ (§19). This is a relation between numerical measures (§2), but it implies certain relations between the units used in measuring these

quantities; both sides must be of the same dimensions in fundamental units, or they could not be equal. Hence, if we denote the unit of velocity by $[V]$, $[L]=[V][T]$ or $[V]=[L][T]^{-1}$. Thus velocity is of $+1$ dimension in length and -1 dimension in time.

153. Dimensions of Acceleration.—Consider any relation between acceleration length and time, such as $s=\frac{1}{2}at^2$. From this by the line of reasoning explained in the last section we derive at once $[L]=[A][T]^2$. Hence $[A]=[L][T]^{-2}$. The sign of equality in such expressions denotes equality of dimensions. Constant numerical factors (such as the $\frac{1}{2}$ above) are of zero dimensions, that is, they do not change when we change the fundamental units.

154. Other Derived Units.—The above examples sufficiently explain the method by which the following table is derived.

TABLE OF DERIVED UNITS USED IN MECHANICS

Quantity.	Relation of Numerics.	Dimensions of Units.	Name of Unit in c.g.s. System.
Linear velocity, v	$s=vt$	$[L][T]^{-1}$	
Linear acceleration, a	$s=\frac{1}{2}at^2$	$[L][T]^{-2}$	
Angular velocity, ω	$\phi=\omega t$	$[T]^{-1}$	
Angular acceleration, α	$\phi=\frac{1}{2}\alpha t^2$	$[T]^{-2}$	
Force, F	$F=ma$	$[L][T]^{-2}[M]$	dyne
Moment of Force, L	$L=Fp$	$[L]^2[T]^{-2}[M]$	
Moment of Inertia, I	$I=mr^2$	$[L]^2[M]$	
Work, W	$W=Fs$	$[L]^2[T]^{-2}[M]$	erg
Kinetic Energy, E.	$K.E.=\frac{1}{2}mv^2$	$[L]^2[T]^{-2}[M]$	erg
Potential Energy	$P.E.=Fs$	$[L]^2[T]^{-2}[M]$	erg

155. Examples of Use of Dimensional Relations.—Where a derived unit has no particular name its dimensional formula is a sufficient name. Thus the unit of acceleration has no special name and 10 units of acceleration in the C.G.S. system is written

$$10\frac{\text{cm.}}{\text{sec.}^2}$$

A frequent use of dimensional relations is in changing the measure of a quantity from one absolute system of units to another. For example, the acceleration of gravity is $980\ \frac{\text{cm.}}{\text{sec.}^2}$, what is it in $\frac{\text{meter}}{\text{min.}^2}$? Suppose it is x. Then

$$x\frac{\text{meter}}{\text{min.}^2}=980\frac{\text{cm.}}{\text{sec.}^2}\ \therefore x=980\frac{\text{cm.}}{\text{m.}}\left(\frac{\text{min.}}{\text{sec.}}\right)^2=35{,}280.$$

Another use of these relations is in testing the accuracy of complicated formulas. The two sides of the equation must be of the same dimensions or they could not stand for the same thing.

PROPERTIES OF MATTER

Constitution of Matter

156. In the preceding chapters on the principles of Mechanics, we have had (with slight exceptions) to consider matter from but one point of view, namely, its inertia. The forces that the particles of a body exert on one another did not need to be considered, for they cancelled out when the action of the body as a whole was considered.

We shall now consider other important properties of matter, especially those which depend on the force between particles. It will be seen that the connections between these properties are not so well understood as the relations between the quantities studied in mechanics. This is chiefly because the ultimate particles of a body are so small that they cannot be studied separately. In fact we can only infer their existence and relations from the properties they exhibit in the large groups which we call bodies.

157. The Three States of Matter.—Following popular language we classify bodies as solids and fluids. The characteristic of a *solid* is that it has a definite shape which it does not readily relinquish, while a *fluid* flows easily or changes its shape in response to the smallest influence. (It will be seen later that the distinction is not quite definite, that some bodies lie on the borderland between the two classes.) The particles of a solid are held in (practically) fixed positions by the forces between them, but each particle has a freedom to vibrate about its mean position (see §161).

Fluids are divided into *liquids* and *gases*. The peculiarity of a liquid is that, while it readily flows, it has a definite volume which it does not readily change. A gas yields to the smallest force exerted to change its volume, in other words, it has no definite volume of its own, but takes the volume of the containing vessel however large. (This distinction also is only general.) The particles of a liquid are close together and attract each other with powerful forces. These forces react strongly against outside forces that tend to change the mean distance between the particles, but they are such as to permit sliding motions of the particles. The particles of a gas are practically separate bodies flying in space and exerting no appreciable forces on one another except at impact of particle on particle.

158. Elements and Compounds.—In innumerable cases two or more substances coalesce to form a new substance that may be so distinct in all its properties that nothing apparently remains to suggest the constituents from which it was formed. Thus two substances, oxygen and hydrogen, gaseous under ordinary conditions, combine to form a liquid, water. Harmless substances may on combination form deadly poisons or explosives. Substances that may be made from constituents which have properties distinct from the resultant are called *compounds*.

Conversely, compounds may be divided up into constituents differing widely from the original substance and these constituents may be themselves capable of being resolved into other constituents. But there are many substances which have not as yet been resolved into constituents and such are called *elements*. Of these there are about 80 known.

159. Molecules and Atoms.—Many facts, chiefly such as are more closely studied in chemistry, justify the belief that (1) an element consists of very small particles called *atoms*, (2) all the atoms in one elementary body are identical in size and other properties, but different from those of any other elementary body, (3) these atoms are combined in similar groups called *molecules* (in some substances the atom and the molecule are identical). It is also believed that a compound consists of molecules and that each molecule consists of two or more atoms of the constituents of the compound. There is also much reason to believe that in many substances, especially liquids and solids, molecules are frequently combined to form groups or molecular aggregates of two or more molecules each.

Molecules and atoms are extremely small and will probably never be separately visible, however much optical instruments may be improved. Thus in a cubic centimeter of a gas under ordinary conditions there are about 4×10^{19} molecules.

There is good reason to believe that atoms contain still smaller parts called electrons, which may pass from atom to atom and are sometimes entirely separated from atoms. The properties of these explain many of the phenomena of light and electricity.

160. Intermolecular Forces.—It is evident from the great forces necessary to pull a solid body apart that there are comparatively great forces between particles; but the ease with

which a brittle body falls apart when a slight crack appears shows that the forces are only appreciable when the attracting particles are very close together. The latter point is also shown by the fact that a body reduced to powder, *e.g.*, the graphite of which lead pencils are made, can only be changed back into a compact solid by intense pressure.

Roughly speaking, it may be said that the force of molecular attraction in water is inappreciable at distances greater than about .00000015 cm. The magnitude and the range of the intermolecular forces are, of course, different for different substances, and the characteristic properties of different substances probably depend on these differences.

161. Kinetic Theory of Matter.—There is very strong evidence that the particles of which bodies are made up are in no case at rest. Thus two different gases contained in two different vessels mix with great rapidity when the vessels are put in communication. This process is called *diffusion*. Liquids will also diffuse into one another (except non-miscible liquids like oil and water), though much more slowly than gases, because of the greater closeness of the particles and the frequent changes of direction of motion of a particle produced by impact on other particles. Even many solids show by diffusion that their particles are not at rest; thus when a small block of gold is placed on a block of lead with planed surfaces in close contact, after the lapse of some weeks it is possible to detect particles of gold which have wandered into the lead and *vice versa*. There are many other reasons for believing that the particles of matter are in all cases in motion. This hypothesis is called the hypothesis of the *kinetic constitution of matter.*

162. Density and Specific Gravity.—The *density* of a body is its *mass per unit volume.* If the masses of all equal volumes of the body are the same, the density is *uniform* and equal to the mass in any unit of volume. If the masses of equal volumes are not the same, the density is not uniform. The *mean density* in any particular volume of the body is the mass in that volume divided by the volume. The *density at a point* is the mean density in a small volume enclosing the point when the volume is supposed to be decreased without limit.

The measure of the density of a body depends, of course, on the units of mass and volume employed. If the c.g.s. system is

employed, density is the number of gms. per c.c. In this system the density of water at 4°C. is very nearly unity, since the gram was originally intended to be the mass of 1 c.c. of water at 4°C. In the British system the density of a body is the number of lbs. per cu. ft. of a body. In this system the density of water is 62.4, since that is the number of lbs. in a cu. ft. of water.

The *specific gravity* of a body is the ratio of its density to that of some standard substance. The standard usually employed is water at 4°C. Thus if D be the density of a body and d that of water at 4°C. the specific gravity of the body is D/d. Now in the c.g.s. system d is very nearly unity. Hence in this system density and specific gravity are numerically equal. But in the British system, since d is 62.4, the specific gravity of a body is its density divided by 62.4.

TABLE OF DENSITIES (GMS. PER CM.^{3})

Aluminium	2.60	Iron (about)	7.60
Brass (about)	8.50	Lead	11.37
Copper	8.92	Platinum	21.50
Gold	19.32	Silver	10.53
Ice	.917	Air at 0° and 1 atmo	.001293
Alcohol at 20°	.789	H " " " " "	.00008988
Ether " "	.715	N " " " " "	.001256
Mercury " "	13.55	O " " " " "	.001430
Sea Water "	1.026		

PROPERTIES OF SOLIDS

163. Homogeneity and Isotropy.—A **homogeneous** body is one which has at all points the same properties, so that small spheres of equal radii cut out of different parts of the body would be identical in properties. Many crystals are nearly perfectly homogeneous, and so, too, is good glass, such as plate glass or the glass of lenses. Many other bodies are approximately homogeneous, for example, most metals, wood, stones, etc.

An **isotropic** body is one which has at any point the same properties in all directions, so that if at any point a sphere were cut out there would be nothing in the properties of the sphere to indicate the original direction of any diameter. All liquids and

gases are isotropic under ordinary conditions but many substances, such as crystals, woods and drawn metals, are distinctly non-isotropic.

164. Elasticity.—When the shape or volume of a solid is changed by the application of some force, there is in most cases a tendency to return to the original shape or volume when this force is removed. This tendency to recover from distortion is called *elasticity*. It is one of the most important properties of a solid, since the usefulness of many bodies such as springs, musical instruments, etc., depends on the extent to which they possess this property. It is, therefore, a property that has been very extensively studied.

165. Strain.—*Any change of shape or of volume or of both is called a strain.* Thus the bending of a beam, the twisting of a rod, the compression of a liquid or a gas into a smaller volume are strains. The term strain is a geometrical one and its definition contains no reference to force or energy, although, as we shall see, force and energy are present when a body is in a state of strain.

A strain that consists in *a change of shape only* without any change of volume is called a *shear*. The strain of a moderately twisted wire or rod is a shear.

A strain that consists in *a change of volume* only without any change of shape has not received any special name, but we may for brevity call it *a volume-strain*. Such, for example, is the strain of a sphere of cork or of any isotropic body when placed in a fluid which is subjected to great pressure in a closed vessel.

While for simplicity we have first enumerated strains in which either volume or shape alone changes, strains which involve changes of both are more common. Thus the stretching of a wire, the compression of a pillar, the bending of a beam, etc., are strains of both volume and shape. A body is said to be homogeneously strained, or the strain is described as *homogeneous*, when the nature and magnitude of the strain is the same at all points in the body. Thus, when a wire is stretched or a rod compressed and when a liquid or gas is subjected to pressure, the strain is homogeneous. But, when a wire or rod is twisted, the strain is greatest at the surface and least at the center, and, when a beam is

bent, there is a stretching on the convex side and a compression on the concave side and the strain is heterogeneous.

166. Stress.—When a body is in a state of strain owing to the action of external forces on it, there are internal forces between contiguous parts of the body in addition to whatever internal forces there may have been before the strain occurred. If at a point a dividing plane be imagined, the part of the body on one side will act with a certain force on the part on the other side and the latter will react with an equal and opposite force. These two forces together, the action and the reaction, constitute a *stress*. In some cases, as we shall see, the stress is perpendicular to the imaginary dividing plane and in others parallel to it, but in any case *the magnitude of the stress is the force per unit area* of such an imaginary dividing plane.

The terms *homogeneous* and *heterogeneous* apply to stress just as to strain. In many cases, for example in the stretch of a wire by an attached weight, the stress in a body is equal to the external force per unit area that acts on the body and produces the strain, and in such cases we may speak of this external force per unit of area as the stress. In other cases, as, for example, in the bending of a beam by a weight acting at some point, the stress does not bear a simple relation to the external force and we must take care to distinguish them.

167. The Measure of a Strain and of a Stress.—A strain which consists in a change of volume only is measured by *the proportion in which the volume is changed.* If the strain is homogeneous the measure may be taken as the change in unit volume, or if a volume V becomes $(V+v)$ the measure of the strain is v/V. If the strain is not homogeneous the measure of the strain at any particular point is the value of v/V at the point, when V is taken as the volume of an indefinitely small portion surrounding the point. To produce this change of volume force must be applied to the surface of the solid in the form of either pressure or a tension, and inside the body each part will press or pull on each neighboring part. The amount of this pressure or pull per unit area is the measure of the stress.

The measure of a *shear* will be most readily understood by considering the simplest way in which a shear may be produced. Consider, for example, a rectangular block of a firm jelly between

two boards to which it adheres. Let $PQRS$ be one rectangular face and PQ, RS the edges of the boards. Apply to the boards equal and opposite forces parallel to them and to the face $PQRS$. The face $PQRS$ is changed to the form P_1Q_1RS. Each section of the block parallel to the boards moves parallel to itself a distance proportional to its distance from RS. Each of the right angles of $PQRS$ is changed by the same amount, say θ, and this change is the measure of the shear. When θ is small, as it is in most practical cases, the magnitude of the angle θ in radian measurement is $P_1P \div PS$, or taking PS equal to unity, *the relative displacement of two planes unit distance apart.*

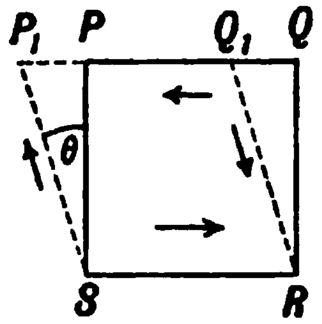


FIG. 79.—Shear and shearing stress.

If an imaginary plane be supposed drawn anywhere in the block parallel to the boards, the part on one side of this plane will exert a tangential force on the part on the other side and this force will equal the force applied to the boards. The *magnitude of the force per unit area* is the measure of the shearing stress.

While we have referred only to the forces parallel to PQ and RS, it is clear that the shear cannot be produced without other forces applied to the block. If only the two forces described were applied, the block would not be at rest but in rotation, since the two constitute a couple. The effect is readily perceived when the attempt is made to apply the two opposite forces. It is, in fact, necessary to also apply other forces forming an opposite counterbalancing couple, say along SP_1 and Q_1R. The effect of all four forces is to produce a stretch along RP_1 and a compression along Q_1S and the proportional stretch is equal to the proportional compression, since there is no change of volume.

168. Hooke's Law.—When any body is strained beyond a certain amount and then released, it fails to return completely to its original form and volume or it retains a permanent set. The largest strain of any kind which a body may undergo and still completely recover from when released is called *the limit of elasticity* for that form of strain, and the corresponding stress is called the *limiting stress.* The limit of elasticity is, of course, widely different for different substances. Thus, rubber may be greatly extended and yet recover, while the limit for glass and

ivory is very small. (Cases in which the limit is somewhat indefinite will be considered later.)

Within the limit of elasticity a simple law, first stated by Hooke in 1676 and known as *Hooke's law*, holds, namely, "*stress is proportional to strain.*" (Hooke's statement in Latin was "Ut tensio sic vis.") Hooke illustrated his law by various cases of strain, such as the stretching of a spiral spring and of a wire, the bending of a beam, the twisting of a wire and so on.

169. Moduli of Elasticity.—While elasticity has already been defined as the tendency of a body to recover its shape or volume when distorted, the definition is purely qualitative and affords no means of assigning a numerical value to the elasticity of a substance. A quantitative definition of the elasticity of a substance for any form of strain follows from Hooke's law. *The measure or modulus of elasticity is the ratio of the magnitude of the stress to that of the accompanying strain*, this ratio being a constant within the limits of elasticity. As there is a great variety of forms of strain there is a correspondingly large number of moduli of elasticity for any substance; but only a few of these are important enough to be enumerated.

When the strain is one of volume only the elasticity is called *elasticity of volume*. The modulus of elasticity of volume or the *bulk-modulus*, as it is frequently called, is the ratio of the stress, or the pressure per unit area, P, to the change of volume per unit volume. The bulk modulus of a substance is usually denoted by k. Hence, if a volume V undergoes a change of volume v and the stress is P,

$$k=\frac{P}{\frac{v}{V}}=\frac{PV}{v}$$

The reciprocal of the bulk modulus is called the coefficient of *compressibility* of the substance. It means the ratio of the proportional compression to the pressure per unit area, or, supposing, the latter to be unity, the coefficient of compressibility is the ratio in which the volume is reduced by unit pressure per unit area.

When the strain is a shear the modulus of elasticity is called the *shear modulus*, or often the *simple rigidity*, and is the ratio of the shearing stress to the shear. Denoting the shearing stress

by T, the shear corresponding to T by θ, and the shear modulus by n,

$$n = \frac{T}{\theta}$$

170. Torsion.—When a wire or rod of homogeneous isotropic material is twisted, we may imagine the whole length divided into transverse slices of equal thickness by planes perpendicular to the axis. Each such slice will be rotated about the axis to an extent proportional to its distance from the fixed end. Moreover, one face of each slice (the one farthest from the fixed end) will be rotated more than the other. Let us now suppose that each slice is very thin, and that it is divided up before twisting into very small cubes (or nearly cubes) by a series of imaginary planes through the axis intersected by concentric cylinders. Thus each cube will have four edges parallel to the axis, four others in the direction of radii, while the remaining four will be short and practically straight arcs of circles. After the twist each cube will have a strain like the cube of jelly in §167. Hence the strain is a shear, but, since the strain of each cube will be proportional to its distance from the axis, the strain is not homogeneous.

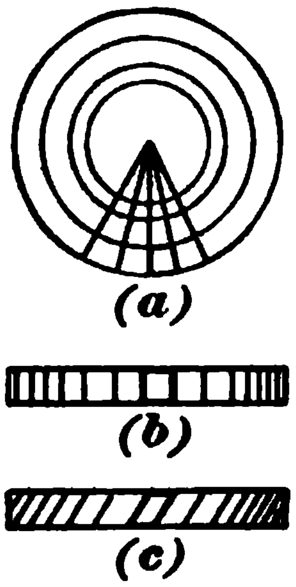


Fig. 80.—Shear of a small cube in a twisted wire.

The *constant of torsion* of a wire has already been defined in §119.

171. Young's Modulus.—A very frequent form of strain is that of a uniform wire or rod which is clamped at one end and is acted on by a longitudinal force at the other end. Such a strain is called a *stretch*. Any short part of the wire is extended in the same proportion as the whole wire. The measure of the stretch is the extension per unit length, or, denoting the unstretched length of the wire by L and the total extension by l, the stretch is l/L. The measure of the stress is the external pull per unit of cross-sectional area. Denoting by F the whole force applied to one end and by a the cross-sectional area of the wire, the pull per unit area anywhere in the rod due to the force F is F/a, which is, therefore, the measure of the stress. Young's modulus, which we may denote by M, is, therefore, $(F/a) \div (l/L)$ or

$$M = \frac{FL}{al}$$

For some common materials the average values of k, n, and M in c.g.s. units, that is, dynes per cm^2 are as follows:

	k.	n.	M.
Copper	17×10^{11}	4×10^{11}	11×10^{11}
Glass	4×10^{11}	2×10^{11}	6×10^{11}
Iron (wrought)	15×10^{11}	7×10^{11}	19×10^{11}
Lead	4×10^{11}	2×10^{11}	1×10^{11}
Steel	17×10^{11}	8×10^{11}	23×10^{11}

172. Volume Changes when a Wire is Stretched.—When a wire or rod is stretched, there is obviously a change of shape in every part of the wire or rod, for the length is increased while the cross-section is decreased. Whether a change of volume also occurs can only be determined by experiment. If the cross-section diminishes in the same proportion as that in which the length increases, there is no change of volume; whereas, if the proportion in which the length increases exceeds that in which the cross-section diminishes, there is an increase of volume. Careful experiment shows that in all cases there is an increase of volume; but in some substances. *e.g.*, india rubber, the change of volume is very small.

173. Flexure.—A very common strain closely related to stretching is that of a plank supported at both ends and carrying a load at the middle, or supported at the middle and loaded at each end, or clamped horizontally at one end and loaded at the other end. A little consideration will make it clear that in these cases we have to do with stretches and shortenings such as those already discussed. If we suppose the plank divided into a large number of longitudinal strips, the strips on the convex side are stretched by the bending, while those on the concave side are shortened. There must, of course, be an intermediate surface where there is neither stretch nor compression and this surface is called the *neutral surface*. The extension or compression of any strip is proportional to its distance from the neutral surface. Thus the strain, while not homogeneous, is everywhere of the nature of an extension or a compression and Young's modulus is the only modulus involved. If a bar of length l, breadth b, and depth d be supported at both ends and be subjected to a perpendicular force F at the middle, the depression produced is $Fl^3/4Mbd^3$.

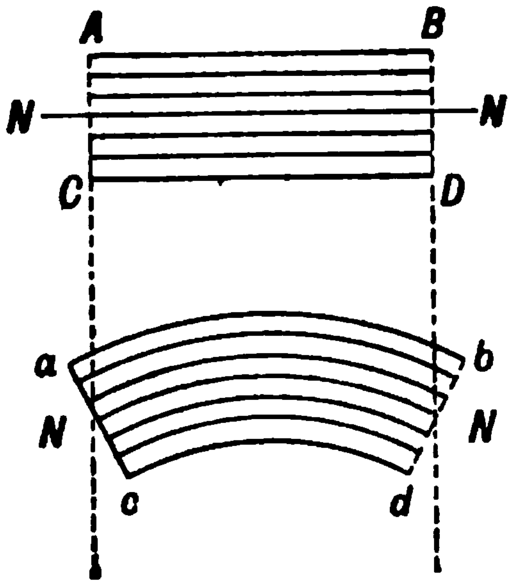


FIG. 81.—Bending of a beam (exaggerated).

parallel to the surface is not changed, while that perpendicular to the surface is changed in the ratio $e:1$. Hence

$$v \sin b = u \sin a, \quad v \cos b = eu \cos a.$$

Hence, by dividing corresponding sides,

$$\tan b = \frac{\tan a}{e}$$

Thus the direction of rebound is more nearly parallel to the surface than that of impact. This is the basis of a method that has been employed for finding e.

176. Loss of Energy on Impact.—The kinetic energy of two smooth spheres before impact and that after impact can be calculated from their masses and velocities. The total kinetic energy of two bodies is less after impact than before (except when e is unity) and other forms of energy, such as heat and sound, are produced.

177. Vibration of Elastic Bodies.—When a body is strained within the limit of elasticity, the internal stresses tend to restore the body to its original condition. When released from the external deforming force the body vibrates, and, since the restoring forces are at each stage proportional to the distortion, the vibrations are simple harmonic vibrations of a constant period. This, for instance, is the case when a rod firmly clamped at one end is bent and released. When the vibrations are sufficiently rapid, as is the case of the prongs of a steel tuning fork, sound is produced, and the ear can test constancy of the period of vibration by the steadiness of the pitch; the vibrations gradually die down, that is, the extent of the maximum strain in each vibration decreases, yet the period remains unchanged, showing that within the limits of vibration the stress is, so far as the delicate sense of hearing can detect, accurately proportional to the strain. A tuning fork can be made of any metal, of wood or other solid substance; and, while the sound may in many cases be weak and short-lived, the steadiness of pitch while it lasts is an excellent proof of Hooke's law.

178. Strain Beyond the Elastic Limit.—As an illustration of what happens when a substance is strained beyond the elastic limit, that is, beyond the range in which Hooke's law holds, we shall consider the stretching of a wire. When a force that

stretches it beyond the limit is applied to it and this force is steadily increased, it elongates in greater proportion for each successive equal increase of the force. As the force is increased, at a certain strain, called the *yield point*, a very rapid increase of strain sets in at some point of the wire, and the strain at that point continues to increase, even if the force is not increased, until at last the specimen "necks in" and breaks. Beyond the yield point the substance flows much like a very viscous liquid. If during this process the force be diminished somewhat, the strain will still continue to increase, but at a diminished rate; and, when the force is diminished sufficiently, the strain ceases to increase before breaking occurs. If at this stage the applied force be removed entirely, the wire will contract somewhat, but a large permanent set will remain. The wire will then act like a different wire with a new elastic limit.

179. Elastic After-effects.—From strain within the elastic limit the strained material completely recovers in time and there is no permanent set; but frequently the immediate recovery on removal of the force is not complete, and there remains a small temporary set from which the material only slowly recovers. This slow recovery from temporary set is called an *elastic after-effect.* It is shown by rubber and glass and other substances which consist of mixtures of diverse molecules; but crystals and quartz threads do not show it.

It is readily demonstrated by clamping both ends of a rubber cord (used for tires of small wheels) and attaching a small mirror to the middle to reflect a beam of light on a scale. Such an arrangement will show a double after-effect due to successive twists in opposite directions.

180. Fatigue of Elasticity.—The vibrations of a torsional pendulum are maintained by the elasticity of the wire; they slowly die away, owing to air resistance and internal friction in the wire. If the pendulum be by some means kept vibrating a long time and then released, the vibrations will die away more rapidly than before, as if the elasticity had become somewhat exhausted by prolonged exercise. This fatigue will persist for a long time but the wire will promptly recover after being heated to about 100°C.

181. Miscellaneous Properties of Solids.—There are many mechanical properties of solids, frequently mentioned, which are not yet defined with

sufficient clearness to make it possible to measure them, but which call for some mention.

A malleable body is one which can be hammered into thin sheets. The most malleable substance is gold, which can be reduced to sheets of gold foil 1/250,000 inch in thickness.

A ductile substance is one which can be drawn out into fine wires. Silver and copper are very ductile; wires less than 1/1000 inch in diameter are readily made from these metals. By heating a substance until it is semi-liquid and then drawing it out, fine threads of substances not ordinarily ductile can be made. Fine tubes and threads of glass are obtained in this way and fine threads of quartz, called quartz fibers, are thus made for use in suspensions of galvanometers and other instruments; they enabled Boys to greatly reduce Cavendish's apparatus (§146).

A plastic substance is one which can be moulded by pressure. Many substances not ordinarily regarded as plastic are so, when subjected to great pressure slowly applied. A stick of sealing wax is ordinarily brittle, but, suspended horizontally on end supports, it will slowly yield to its weight and bend. All metals under enormous shearing stresses become plastic. The impact of a cannon ball on armor-plate will sometimes produce a splash like a stone dropped in water.

A friable substance is one easily reduced to powder by a blow. Glass, diamond and crystals are friable.

Hardness is a term used in different senses. It sometimes means the opposite of plasticity, that is, resistance to change of shape, as when we speak of iron as hard and rubber as soft. Another use of it is to denote power of scratching, as in the mineralogists' scale of hardness, which consists of a series of substances, with diamond at one end and talc at the other, arranged so that each, beginning with diamond, will scratch the following but not the preceding. Any other substance that will scratch one in the list but not the next higher is said to have a hardness between the two.

PROPERTIES OF FLUIDS

182. A **fluid** is distinguished from a solid by the absence of permanent resistance to forces tending to produce a change of shape; that is to say, *the shear modulus of a fluid is zero.* In this respect all fluids agree; they also agree in having weight and inertia. Because of agreement in these respects there are certain properties common to all fluids.

In certain other respects liquids and gases differ considerably. These differences are due to the fact that, while the particles of liquids are comparatively close together and attract one another with very considerable forces, the particles of gases are so far apart that the forces between them are usually negligible (except

at impact). Properties in which liquids and gases differ will therefore, be treated in separate chapters.

183. Direction of Force on the Surface of a Fluid.—When a fluid is at rest, the force acting on its surface must be perpendicular to the surface. This results from the fact that the shear modulus is zero; for, if the force were not at right angles to the surface, it might be resolved into a component perpendicular to the surface and a component along the surface. The latter would produce a sliding motion, or a shear of the liquid near the surface, so that the liquid could not be at rest.

At the surface of contact of a fluid and a solid, for example at any part of the surface of a vessel in which the fluid is contained, the force exerted by the fluid on the solid is at right angles to the common surface. If it were not, the reaction of the solid on the fluid, being equal and opposite to the force of the fluid on the solid, would not be at right angles to the surface of the fluid.

At the free surface of a liquid, that is, where the liquid is in contact with a gas, the pressure between the two must be at right angles to the surface. The force of gravity must also be at right angles to the free surface of a liquid at rest or sliding motion would result. Hence the free surface of a liquid at rest is horizontal unless it is acted on by some other force than gravity and gas pressure, such as surface tension (which we shall consider later) or magnetic force acting on a magnetic liquid.

184. Pressure in a Fluid.—In a fluid there are forces, actions and reactions between contiguous parts of the fluid. These forces are due to several causes. The *weight* of the upper layers of the fluid has to be sustained by the lower layers and a pressure thus results. Force on the surface of the fluid, if it be completely enclosed, produces a pressure in the fluid; this is true not only of a fluid in a vessel which it completely fills, but also of a liquid the free surface of which receives the pressure of the atmosphere or of any gas above the liquid. (Another cause of pressure in a fluid is referred to in §206.)

The total force exerted by a fluid on any surface is called the *thrust* on that surface. The thrust per unit of area at a point on the surface is called the *pressure intensity* or simply the *pressure* at that point. The pressure over a surface may be either uniform,

that is, the same at every point, or variable. When uniform the pressure at any point equals the force on any unit of area; when variable it equals the average pressure, that is, the force on an area divided by the area, when the area is reduced without limit.

Whatever the causes of pressure in a fluid, the *pressure at a point is the same in all directions*, that is to say, if we suppose an imaginary surface to separate the fluid at a point into two parts, the pressure of each part on the other is, as we have already seen, perpendicular to this surface, and it is also the same no matter how the imaginary surface is supposed to be inclined. This is nearly obvious from the mobility of the fluid, but the rigorous proof of the statement is not difficult.

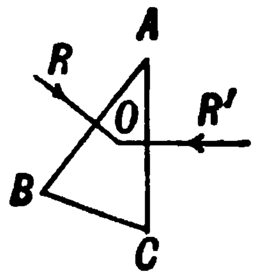


FIG. 84.

Let O be the point considered and let RO and $R'O$ be any two directions through O. Around O suppose a small prism described, and let two of its faces, of which AB and AC are the traces, be perpendicular to RO and $R'O$ respectively; while the third face, of which BC is the trace, is equally inclined to AB and AC, and let the ends of the prism be planes parallel to ABC. The fluid within the prism is at rest and therefore (neglecting its weight for a reason stated later) the thrusts on all its faces form a system of forces in equilibrium. Hence the sum of the components of the forces in the direction BC equals zero. Two only of the thrusts have components in the direction BC, namely, those on AB and AC. Let these be R and R' respectively. They are equally inclined to BC, and if each makes with BC the acute angle θ,

$$R \cos \theta - R' \cos \theta = 0$$

Now the areas of the faces AB and AC are equal: suppose each is a. Cancelling $\cos \theta$ and dividing by a, we get

$$R/a = R'/a$$

If we now suppose the prism to become indefinitely small, R/a becomes the pressure at O in the direction RO and R'/a becomes the pressure at O in the direction $R'O$. Since RO and $R'O$ stand for any directions through O, the pressure is the same in all directions.

As stated above the weight of the prism was neglected. As the prism is diminished without limit, the weight of the liquid in it, which is proportional to the cube of its dimensions, decreases more rapidly than the thrusts, which are proportional to the squares of the dimensions; each time the prism is reduced to one-half in linear dimensions the area of each face is reduced to one-fourth and the weight of the contained liquid is reduced to one-eighth. Hence when the prism is taken small enough the weight becomes negligible compared with the thrusts.

185. Pressure at Different Points in a Fluid.—(1) Let P and Q be two points in a fluid at rest, the positions of the points being such that the straight line PQ is horizontal and wholly in the fluid. Consider the forces acting on a cylinder of the fluid described about PQ as axis. The thrusts on the curved surface of the cylinder have no components in the direction of the axis. Hence, for equilibrium, the thrusts on the ends must be equal and opposite; and, since the ends are of the same area, the average pressures on the ends must be equal. If, now, the radius of the cylinder be supposed indefinitely decreased, the average pressures on the ends become the pressures at P and Q, which must, therefore, be equal. Hence the pressure in any direction at P equals the pressure in any direction at Q.

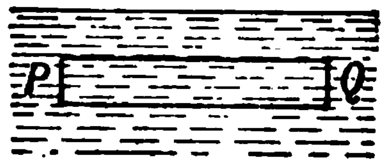


FIG. 85.

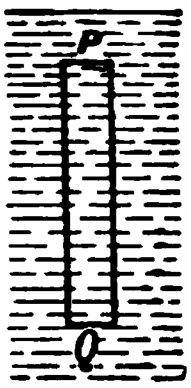


FIG. 86.

(2) Let P and Q be two points in a vertical line wholly in a fluid of density ρ. Consider the forces acting vertically on a cylinder described with PQ as axis and of unit, 1 sq. cm., cross-section. If the depth of Q below P be h cms. the volume of the cylinder will be h c.c., its mass will be $h\rho$ gms. and its weight $h\rho g$ dynes. If p_1 be the pressure in dynes per sq. cm. at P and p_2 that at Q, the thrust downward at P will be p_1 and that upward at Q will be p_2. Hence

$$p_2 - p_1 = h\rho g$$

(3) Let P and Q be any two points in the fluid. No matter what the shape of the containing vessel, P and Q can be connected by a broken line made up of vertical and horizontal steps. Along the zigzag path from P to Q there will be a difference of pressure $h'\rho g$ for each vertical step of length h', while for each horizontal step there will be no change of pressure. Hence the difference of pressure between P and Q will be $g\rho \times$ (the algebraic sum of the vertical steps) or, if the difference of level of P and Q be h, the difference of pressure will be $h\rho g$.

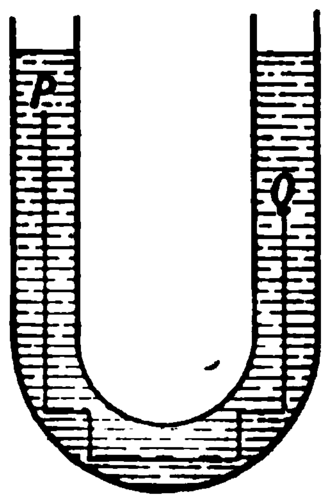


FIG. 87.

186. Pressure in a Gas.—Since the density of a gas is comparatively small, the difference at two points is usually so slight

as to be negligible; but this is not the case if h be very great. Thus in a vessel containing gas the pressure may be regarded as everywhere the same; but the pressure of the air varies greatly as we ascend to great heights in the atmosphere or descend to great depths in a mine.

187. Units Employed in Calculating Fluid Pressure.—In establishing the formula for difference of pressure at different depths in a fluid, namely,

$$p_2 - p_1 = h\rho g$$

it has been supposed that absolute units are employed. If h be in cms., ρ in gms. per c.c. and g in cm. per sec.^{2} (about 980), p_1 and p_2 will be in dynes per sq. cm. A dyne per sq. cm. is sometimes called a *bar*.

When the values of p_1 and p_2 would be inconveniently large in absolute units, other units may be employed. If g be omitted, p_1 and p_2 will be in gms. wt. per sq. cm. and

$$p_2 - p_1 = h\rho$$

This formula may also be used to calculate the pressure in metric tons (1,000,000 gms.) per sq. m. (10,000 sq. cm.) if h be in meters (100 cm.).

When British units are employed the weight of a cylinder of 1 sq. ft. cross-section and h feet in length and of density ρ (lbs. per cu. ft.) is $h\rho$ lbs. Hence if p_1 and p_2 are in lbs. wt. per sq. ft.,

$$p_2 - p_1 = h\rho$$

188. Surface of Contact of Two Fluids.—The surface of contact of two fluids of different densities which are at rest and do not mix is horizontal. This may be deduced from the principle that, for stable equilibrium, the potential energy of a system must be a minimum (§107). If any part of the denser fluid were at a higher level than an equal part of the less dense, the potential energy could be decreased by interchanging the two. Hence, for the potential energy to be a minimum, every part of the denser fluid must be lower than any part of the less dense, that is, the surface of contact must be horizontal with the denser liquid below. Another proof is to suppose that the surface could be inclined, as LM. Let P and Q be two points in the surface.

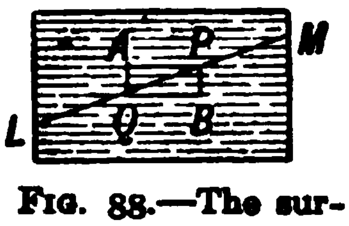


FIG. 88.—The surface of contact of two fluids cannot be inclined as LM is.

Complete a rectangle $AQBP$ with vertical and horizontal sides. The pressure at A would equal that at P and the pressure at Q would equal that at B. The increase of pressure from A to Q would equal the increase from P to B and this could not be when the liquids are of different density.

A particular case of the above is the surface of contact of a liquid with the atmosphere or any gas; it must be horizontal.

In both proofs it has been assumed that gravity is the only force acting on the particles of the fluids; if any other force exist, the surface may not be horizontal. In any case it is at right angles to the *resultant force.*

189. Pascal's Principle.—When a fluid is at rest, the difference of pressure between two points depends only on the difference of level and the density (§185). Hence, if the pressure at any point be increased, there will be an equal increase of pressure at every point (provided the density does not change appreciably) or *pressure is equally transmitted in all directions.* This is Pascal's principle of the transmissibility of pressure.

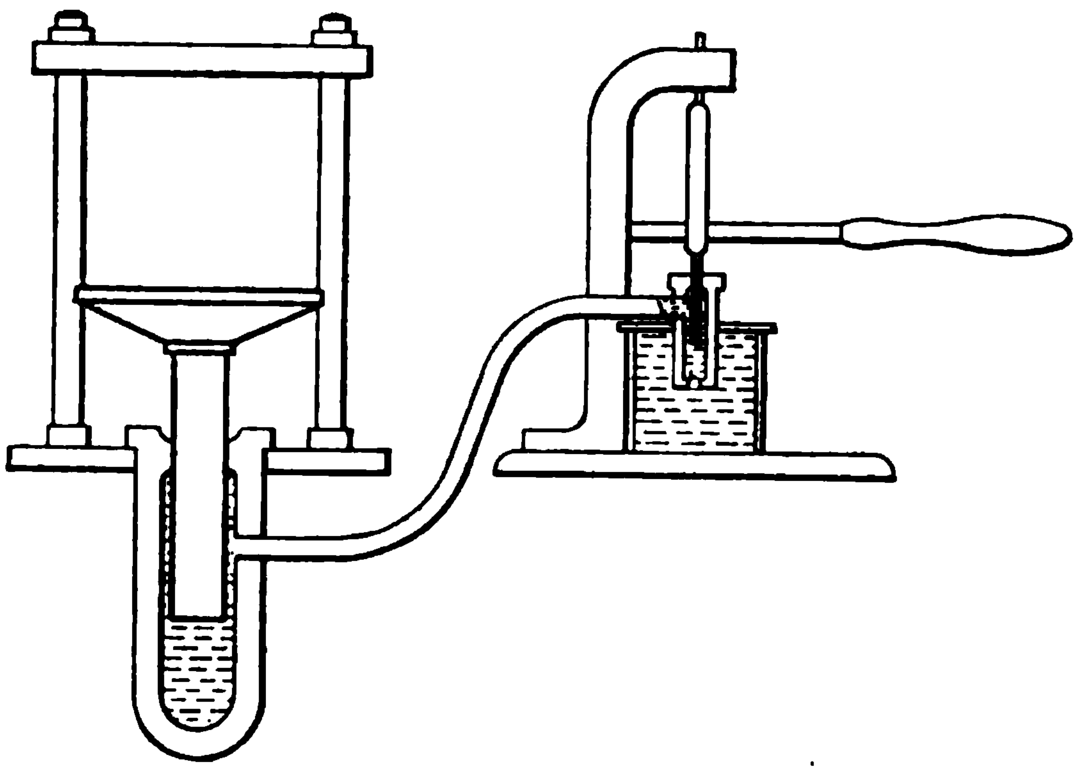


Fig. 89.—Hydraulic press.

Pascal's principle is not rigorously true for a compressible fluid, for pressure will produce a change of density of a compressible fluid. But the compressibility of liquids is so small that the principle is practically true for all liquids. Gases are much more compressible, but their densities under ordinary

circumstances are so small that the pressure in a moderate volume is everywhere practically the same and the principle is practically true for gases also.

190. The Hydraulic Press.—In the hydraulic press Pascal's principle is applied to obtain a great force by the exertion of a relatively small one. It consists of a large cylinder and piston (or plunger) and a small cylinder and piston, the two cylinders being connected by a tube and filled with some liquid. Let the area of the large piston be S and that of the small one s. If the pressure in the liquid is p, the thrusts on the pistons are pS and ps respectively, and these are in the ratio of $S : s$. Hence a small external force applied to the small piston will enable the large piston to exert a relatively great external force. The arrangement is indicated in figure 88; a valve in the connecting tube permits flow from the smaller cylinder toward the larger, but not in the opposite direction.

191. Archimedes' Principle.—When a body is partly or wholly immersed in a fluid at rest, every part of the surface in contact with the fluid is pressed on by the latter, pressure being greater on the parts more deeply immersed. The resultant of all these forces of pressure is an upward force called the *buoyancy* of the body immersed. The direct calculation of this resultant force is difficult except when the body has some simple form, such as a cylinder with its axis vertical; but a simple line of reasoning will show the magnitude and direction of the force.

The pressure on each part of the surface of the body is evidently independent of the material of which the body consists. So let us suppose the body, or as much of it as is immersed, to be replaced by fluid like the surrounding mass. This fluid will experience the pressures that acted on the immersed body and this fluid will be at rest; hence the resultant upward force on it will equal its weight and will act vertically upward through its center of gravity. It follows that *a body wholly or partly immersed in a fluid is buoyed up with a force which is equal to the weight of the volume of the fluid which the body displaces* and which acts vertically upward through the center of gavity of the fluid before its displacement. This point through which the force of buoyancy acts is called the *center of buoyancy*.

Since the weight, in dynes, of the fluid displaced equals the

product of the volume (which equals the volume immersed) its density and g,

$$Buoyancy = Volume\ immersed \times density\ of\ fluid \times g$$

Buoyancy is to be treated as any other force that acts on a body and either causes motion or helps to produce equilibruim. If a body of mass M, wholly or partly immersed in a fluid, be sustained partly by buoyancy and partly by another upward force F, then, in absolute units,

$$F + V\rho g = Mg$$

where V is the *volume immersed* and ρ is the *density of the fluid*. When gravitational units are employed g must be omitted.

192. Fluids in Motion.—While the calculation of the motion of a rigid solid body is comparatively simple, owing to the fact that we may treat a solid as a whole without regard to the actions between its parts, the discussion of the motion of a fluid is rendered difficult by the readiness with which any part of the fluid changes its shape, and we cannot, therefore, without the use of advanced mathematics, treat of any except a very few and simple cases of the motion of fluids.

When a fluid moves either in an open stream or in a closed pipe, the continual change of shape of each part is opposed by internal friction between these parts, and to maintain the motion some external force must be applied to the fluid. The most common causes of motions of fluids are gravity, as in the case of a river, pressure applied to some part of the boundary of the fluid, as in the case of water pumped through a system of piping, and the motion of solids in contact with the fluid, as in the case of a fan.

193. Flow in Pipes.—When the pressure on a fluid in a horizontal tube is greater at one end than at the other, a flow ensues. When the pressure is first applied, the motion begins with an acceleration, but after a time, if the pressure at the ends are kept constant and the supply of fluid is maintained, a steady state of motion ensues, so that at each part of the tube the motion is constant. The simplest case is when the tube is of constant cross-section and the fluid is pratically incompressible, that is, a liquid. In this case the volume of fluid passing all cross-sections of the

pipe is the same throughout, and the rate of flow is, therefore, the same at each cross-section. The motion is from places of higher pressure to places of lower pressure. If, however, the fluid be compressible, while the motion at a point remains steady and the mass that passes every cross-section is necessarily the same as that which enters the pipe, the volume of flow is variable; for where the pressure is greater the fluid is compressed into a smaller volume, and where the pressure is less the fluid is not so much compressed. Thus the speed of the fluid particles is on the whole greater in the parts of the pipe where the pressure is less, that is, the further along the stream we go.

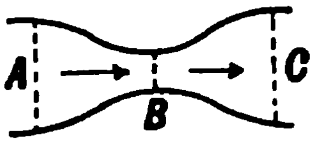


FIG. 90.

When a liquid flows through a tube of variable section (Fig. 90), the pressure at the ends being constant, the mass that passes each cross-section is the same, but the rate of motion of the particles increases as the stream comes to a contraction of the tube, and decreases again as the stream comes to an expansion of the tube. Now an increase of velocity or an acceleration necessarily means a smaller pressure ahead than behind, and a decrease of velocity necessarily means a larger pressure ahead than behind. Thus in a contraction (or "throat") the pressure is smaller than immediately before or behind, the amount of difference being dependent on the rate of flow through the tube and the cross-section at the throat and at either side. This principle is the basis of the *Venturi meter* for gauging the flow of water in pipes. The same principle is employed in the *aspirator* (Fig. 91), a form of air-pump attached to a water faucet. The water is forced through a contraction in a tube and produces suction in a side tube which is connected to the vessel to be exhausted. Similar considerations apply to the flow of gas through a pipe of variable cross-section, but this case is complicated by the changes of volume due to changes of pressure.

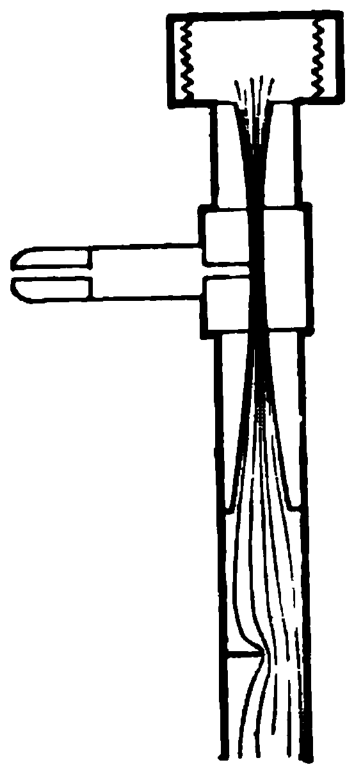


FIG. 91.

194. Illustrations of the Above.—Fig. 92 represents (in section) a glass tube that passes tightly through a wide cork and a second cork

through the center of which a pin is stuck. When air is blown through the tube the lower cork is not repelled but is attracted (the pin prevents side motion). The air increases its speed to pass through the small space between the corks, hence its pressure diminishes and atmospheric pressure pushes the corks together. Various other pieces of apparatus, such as the atomiser, the ball nozzle and the injector, act on the same principle.

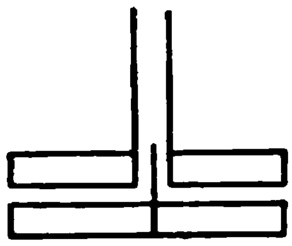


FIG. 92.

The curvature of the path of a rotating base ball or tennis ball is due to a difference of pressure on the opposite sides of the ball. Suppose the ball had no translatory motion but had a motion of rotation, while a current of air blew on it as indicated in figure 93. The rotating ball would carry air around with it. At A the two air motions would conspire and at B they would be in opposition. Hence the velocity of the air-particles would be greater at A than at B and the pressure at B would, therefore, be greater than that at A, the result being a force on the ball in the direction BA. The same result follows when the ball is moving through air otherwise at rest, and the path curves toward the side of less pressure. The motion of the ball can also be explained by considering that the impacts between the ball and the air particles are necessarily more violent on the side B than on the side A.

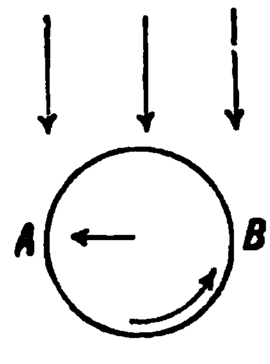


FIG. 93.—"Curve" of a base ball.

195. Work Done by a Piston.—When a piston of area a moves a small distance d along a cylinder against a pressure P (per unit area), it exerts a force Pa through a distance d, and therefore does work Pad. Since ad is the volume, say ΔV, of the small part of the cylinder through which the piston has moved, the work done is $P. \Delta V$.

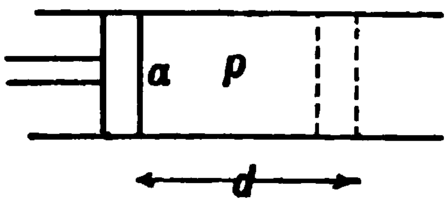


FIG. 94.—Work by a piston.

If the pressure is not constant, as, for example, where a gas in a closed vessel is being compressed, the whole work will be the sum of products $P. \Delta V$, where P must be given its proper value for each successive change of volume ΔV. We may also use a graphical method as in §22 and §56. In the present case (Fig. 95) each abscissa will represent the volume at some moment and the corresponding ordinate will represent the pressure at that moment. The area will represent the whole work.

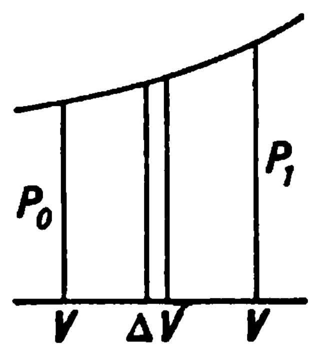


FIG. 95.

Conversely, an expanding fluid does work equal to the sum of

$P.\Delta V$, where P is the pressure and ΔV an increment of volume. If P is constant and the whole change of volume is V, the work done is PV.

196. Viscosity.—A fluid offers no permanent resistance to forces tending to change its shape; it yields steadily to the smallest deforming force. But the rate of yielding is different for different fluids, that is, different fluids offer different *transient* resistances to deformation. This transient resistance is called internal friction or *viscosity*. Thus a very viscous liquid such as glycerine or tar flows much more slowly through a tube or down an incline than water does, and such a flow consists in a continuous change of shape of each part of the liquid. The internal forces are what we have called stresses, and, since the strain is a change of shape only, the stress must be a shearing stress.

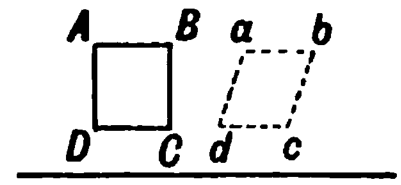


FIG. 96.—Shearing of a fluid.

Consider, as an example, a stream (Fig. 96) flowing down a very gentle incline under the force of gravity. The motion is greater near the surface than at the bottom. A small cube $ABCD$ with sides vertical and horizontal will, by the motion, be changed into the form *abcd*. The liquid above AB exerts a force in the direction AB, on the upper face of the cube, and the liquid below CD exerts a resisting force on the face CD in the direction CD. These two forces constitute a shearing stress. A similar description applies to a small cube of a liquid flowing in any manner whatever. Very extensive experiments have shown that *the ratio of the shearing stress to the rate of shear is a constant for any one fluid*, the value of the constant being different for different fluids. This is the fundamental and very simple law of fluid friction.

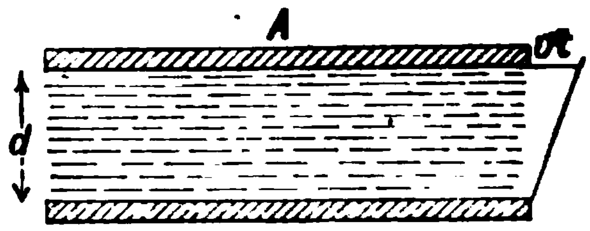


FIG. 97.

The constant ratio of the shearing stress in a fluid to its rate of shear is called the coefficient of viscosity of that fluid.

A concrete case will make this definition clearer and will lead to another way of stating the definition. Suppose (Fig. 97) that the space between two large parallel plates is filled with the fluid under consideration and let one plate be moving parallel to the other with a velocity v. Experiment (as stated later) shows

that the fluid in contact with the plates does not slip along the faces of the plates but adheres to them. The moving plate will in a very short time t move a distance vt, and, if the distance between the plates be d, the shear produced in the time t will be vt/d. Hence the *rate of shear* is v/d. If the area of each plate be A and the force applied to move the upper plate be F, *the shearing stress* will be F/A. Hence, denoting the coefficient of viscosity by μ we have

$$\mu = \frac{F/A}{v/d}$$

$$F = \mu \frac{Av}{d}$$

If, now, we suppose A, v and d to be each unity, F will be equal to μ. Hence we have the following definition of μ: *The coefficient of viscosity of a fluid is the tangential force on unit of area of either of two horizontal planes at the unit distance apart, one of which is fixed while the other moves with the unit velocity, the space between them being filled with the viscous material.*

197. Measurement of Coefficients of Viscosity.—The most common method of finding μ is by measuring the flow of the fluid through a tube of very small bore (or so-called capillary tube). The motion of the fluid in such a case (provided the velocity does not exceed a certain magnitude) is analogous to the slipping of the tubes of a small pocket telescope through one another. If we imagine the fluid divided into a very large number of thin cylindrical shells, the motion consists in the slipping of shell through shell; hence the resistance encountered is internal friction or viscosity. Let p be the difference of pressure per unit area at the two ends of the tube (supposed horizontal), l the length of the tube, and r its radius. It has been shown theoretically and experimentally that, when the fluid is a liquid, the volume that flows out of the tube in unit time is

$$V = \frac{p\pi r^4}{8l\mu}$$

This formula also applies to a gas if p be very small, but if p be large the formula must be modified to allow for the compressibility of the gas. In the theoretical proof of the above formula

it is assumed that no slipping of the fluid along the surface of the tube takes place, and the agreement of theory and experiment confirms this assumption.

The following are some values of μ in c.g.s. units at 20°C.

Alcohol	0.0011	Water	0.010
Ether	0.0026	Glycerin	8.0

198. The Explanation of Viscosity.—Viscous resistance to fluid motion resembles friction between solids in certain respects, and in other respects the two are very different. Both are forces that appear only as resistances to relative motion; they are, therefore, non-conservative forces and energy spent in doing work against them is changed into heat. But, while the friction between solids is, through a considerable range of velocity, independent of the velocity, the resistance due to viscosity is exactly proportional to velocity through the widest range in which experimental tests have been made. This points to a fundamental difference in the nature of the resistance in the two cases.

There are many strong reasons for believing that the particles of fluid are in rapid motion and are not, like the particles of solids, confined to more or less definite positions. If now we imagine two layers of a fluid in relative motion, so that one is passing another, like one railway train passing a second, it is evident that particles from each layer must be continually crossing the boundary into the other layer. The particles of the more rapidly moving layer that cross the boundary carry their larger momentum with them and thus produce a gradual increase of the velocity of the second layer. At the same time particles of the latter layer penetrate into the former and by taking up momentum diminish the velocity of that layer. The result, on the whole, is a tendency of the two layers to come to the same velocity, and this is exactly what we mean by a resistance to relative motion. In the case of gases this explanation may be regarded as fully established; for the formulas to which it leads by mathematical methods are verified by experiment. It has not yet been found possible to work out the mathematical results in the case of liquids, but there is no reason to doubt that the explanation is equally applicable to the latter.

Liquids

199. Compressibility of Liquids.—While the shear-modulus of any liquid is zero the bulk-modulus is usually large, that is, the pressure on a liquid must be greatly increased to produce much diminution of volume. The coefficient of compressibility of a liquid (§169) is therefore small. Measurements of the compressibilities of liquid are made by subjecting the liquids to great

pressures in a vessel called a piezometer and noting the resulting diminution of volume.

The following table gives the compressibilities of some liquids. Each number is the proportion by which the volume of the liquid is decreased when the pressure on it is increased by one atmosphere.

Alcohol	0.0000828	Mercury	0.0000033
Ether	0.0001156	Water	0.0000489

200. Hydrometers.—A hydrometer is an instrument for finding the density of liquids; some hydrometers may also be used to find the density of solids. The action of most hydrometers depends on Archimedes' principle. Some hydrometers sink to different depths in different liquids and thus indicate the densities of the liquids; these are called hydrometers of *variable immersion*. Other hydrometers are used with different weights added to the

Fig. 98.—Common hydrometer.

Fig. 99.—Nicholson's hydrometer.

weight of the instrument so that they are always immersed to the same depth; these are called hydrometers of *constant immersion*.

The common hydrometer is one of variable immersion. It is a glass tube with an enlargement in the middle and weighted at the lower end with mercury so that it will float in stable equilibrium. Inside the tube is a scale which indicates the density of the liquid by the depth to which the tube is immersed (Fig. 98).

The best known hydrometer of constant immersion (Fig. 99) is Nicholson's Hydrometer. It consists of a hollow cylindrical body (of metal or glass), to one end of which a somewhat heavy basket B is attached, while at the other end there is a stem S which carries a scale pan C for weights. On the stem there is a mark indicating the depth to which the hydrometer is to be immersed. Let W be the weight of the hydrometer and let w be the weight that must be placed on the pan to make the instrument sink to the mark in water of density ρ, and w' the weight on C required when the hydrometer is in a liquid density ρ'. The volume of liquid displaced in both cases is the same. Hence, by Archimedes' principle, the weights of equal volumes of the second liquid and of water are $W+w'$ and $W+w$, and their ratio is the ratio of the densities.

This hydrometer may also be used to find the density of a small solid. When so used the instrument is in reality a balance for weighing the solid in air and then in some liquid of known density. The body is first placed on C. The weight required on C to sink the hydrometer to the mark on the stem will be less than w' by the weight of the body. This gives the weight of the body in air. The body is then placed in B and its apparent weight when immersed is found in the same way. The ratio of the weight of the body to its apparent loss of weight when immersed, which equals the weight of an equal volume of liquid, gives the specific gravity of the body relatively to the liquid.

201. Stability of Flotation.—A body floating at rest on the surface of a liquid is in equilibrium under the action of its weight acting vertically downward through the center of gravity, G, and

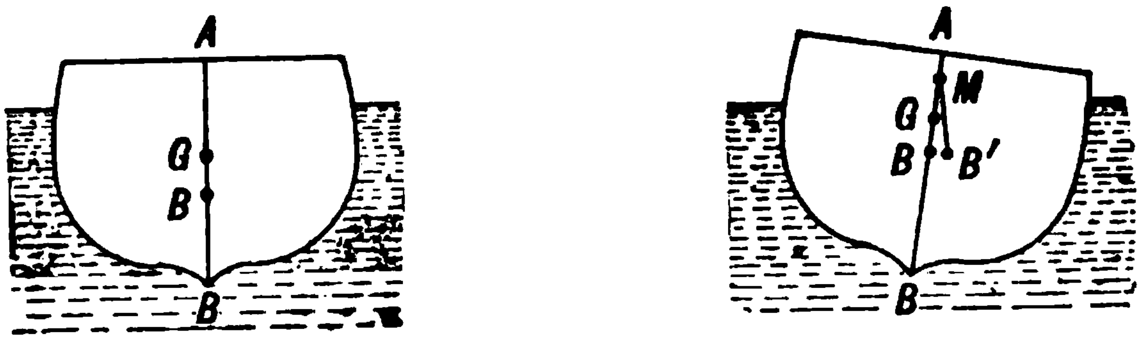


FIG. 100.—Stable equilibrium of a vessel.

the resultant upward pressure of the liquid acting through the center of buoyancy, B. Hence the two forces are equal and act in opposite directions in the vertical line BG. Suppose the body to rotate slightly about an axis perpendicular to the plane represented in the figure. The form of the volume of water displaced is now different (unless the body be spherical or cylindrical), and the center of buoyancy is at some point B' not in the vertical line

through G. Hence the forces now acting on the body constitute a couple, and, if the couple tends to right the body, the equilibrium is stable; if not, it is unstable.

The simplest case to consider is when the body is symmetrical, or very nearly so, on opposite sides of the plane through B and G perpendicular to the axis of rotation; for in this case B' is in this plane. Let a vertical through B' cut BG in M. For small rotations the position of M on BG is very nearly independent of the magnitude of the rotation. M is called the *metacenter* of the body. The position of M can usually be calculated by mathematical methods. If M is above G it is evident that the couple tends to right the body and the equilibrium is stable; if M is below G the couple tends to displace the body further and the equilibrium is unstable. Hence the danger of taking the whole cargo out of a vessel without putting in ballast and the risk of upsetting when several people stand up at once in a small boat. A ship has one metacenter for rolling and another for pitching. In general the vessel is not quite of the symmetrical form assumed above and the problem of stability is more complicated. (Article "Shipbuilding," Ency. Britt., 11th ed.)

202. Energy of a Moving Stream.—When liquid flows steadily through a pipe of varying cross-section, the total energy in the space between any two sections A and B remains constant. When a volume V flows in through A, an equal volume flows out through B. Let the pressure at A and B respectively be p_1 and p_2, and the velocities v_1 and v_2 respectively. Let ρ be the density of the liquid. When the volume V flows in through A it carries kinetic energy $\frac{1}{2}V\rho v_1^2$ into the space between A and B and in the same time the volume V flows out through B carrying energy $\frac{1}{2}V\rho v_2^2$. There is thus a gain of kinetic energy $\frac{1}{2}V\rho(v_1^2-v_2^2)$.

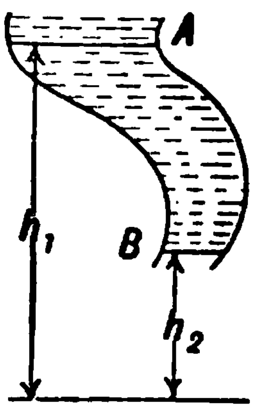


FIG. 101.

Now the liquid above A acts like a piston in forcing liquid into the space between A and B, and it thus does work p_1V which goes to increase the energy between A and B; and in the same time the liquid between A and B does work p_2V in forcing liquid out through B. From this cause there is an increase of energy p_1V-p_2V between A and B. If between A and B there is a fall of level from h_1 to h_2, the liquid which flows in at A will have a greater amount of potential energy than that which flows out at B, and there will, therefore, be an increase of potential energy of $V\rho g\ (h_1-h_2)$ between A and B. But the total energy between A and B remains constant. Hence

$$\frac{1}{2}V\rho(v_1^2-v_2^2)+(p_1V-p_2V)+V\rho g(h_1-h_2)=0$$

or

$$p_1+g\rho h_1+\frac{1}{2}\rho v_1^2=p_2+g\rho h_2+\frac{1}{2}\rho v_2^2=\text{a constant}$$

This is Bernoulli's theorem. It is of great importance in hydraulics.

203. Outflow from an Orifice—Torricelli's Theorem.—When an orifice is opened in a side of a vessel containing liquid at greater than atmospheric pressure, the liquid is forced outward. The simplest way of finding the velocity of the escaping liquid is by an application of the principle of the conservation of energy.

A small mass m of liquid escaping with velocity v has $\frac{1}{2}mv^2$ units of kinetic energy. If no liquid has been added to the vessel during the escape of the mass m, the potential energy of the liquid in the vessel must have diminished by an amount equal to $\frac{1}{2}mv^2$. The mass m was really removed from the part of the liquid near the orifice, but the change of the state of the liquid in the vessel is the same as if the mass m had been removed from the surface; and the change of total potential energy of the liquid in the vessel and of the escaping liquid is the same as if a mass m had been lowered from the surface to the depth of the orifice. Hence, denoting the depth of the orifice below the surface by h, the loss of potential energy is mgh and, therefore,

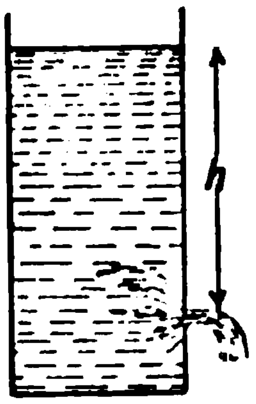


FIG. 102.

$$\frac{1}{2}mv^2 = mgh$$

and

$$v = \sqrt{2gh}$$

Thus *the velocity of escape is the same as if the escaping liquid had fallen freely through the distance of the orifice below the surface.* This is known as Torricelli's Theorem. It was first stated by a pupil of Galileo named Torricelli, who also discovered the principle of the barometer.

Torricelli's Theorem may also be deduced from Bernoulli's theorem, but we shall leave the deduction as an exercise for the reader.

The above theorem relates only to the velocity of the particles as they leave the orifice. It does not enable us at once to calculate the volume that escapes in a given time; for the cross-section of the jet contracts for a short distance after leaving the vessel, and at a certain point reaches a minimum called the *vena contracta* (or contracted vein) beyond which it expands. If the area a of the cross-section of the *vena contracta* is found, the volume per second that escapes is av. The ratio of a to the area of the orifice depends on the velocity of escape and can be changed by inserting a tube (or *ajutage*) through the orifice.

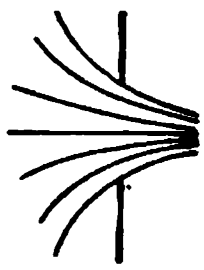


FIG. 103.

204. Pressure Exerted by a Stream.—When a stream of liquid meets an obstacle and is arrested, it gives up its momentum to the obstacle, that is, it exerts a force on the obstacle. The pressure thus produced can be calculated from the velocity of the water and the amount of water that impinges per second on this obstacle. On this is founded a method of measuring the velocity of a stream (Pitot's tube). A tube bent at right angles is placed in the stream so that one arm points horizontally up stream and the other vertically upward. If the water were at rest, the liquid would rise in the vertical arm to the height of the surface of the water, but the pressure of the stream raises it higher and from this additional height the velocity of the stream can be deduced.

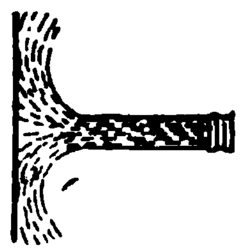


FIG. 104.

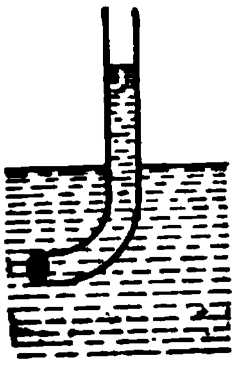


FIG. 105.—Pitot tube for finding velocity of stream.

In Pitot's tube the case of §203 is reversed. Let the rise of level be h. Suppose the liquid to be continually removed at the level h. In any time the total decrease of kinetic energy will be $\frac{1}{2}mv^2$ and the total increase of potential mgh. Hence $v^2=2gh$ (approx.) but a correction factor (slightly greater than unity) is necessary because the tube disturbs the uniform flow of the stream.

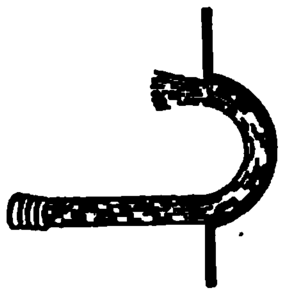


FIG. 106.

When a jet impinges on an obstacle and flows off laterally, the pressure exerted is that due to the loss of the momentum of the liquid. If this obstacle is curved so that the motion of the liquid is reversed, the water is given an equal momentum in the opposite direction and the force exerted on the obstacle is doubled. This principle is taken advantage of in the construction of water-wheels.

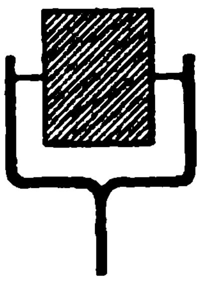


FIG. 107.—A disk swept through the air turns perpendicular to the direction of motion.

When a stream strikes an obstacle obliquely, it is partly arrested and then flows down along the surface of the obstacle. Thus the side of the obstacle farther up stream receives more momentum than the lower side and so tends to turn more nearly perpendicular to the stream. A floating log free only to swing about its middle point sets itself across the stream. A leaf falling from a tree tends to take a horizontal position. The effect is readily illustrated by sweeping through the air a

square disk of cardboard which is connected by short threads to a wire frame (Fig. 107).

205. The Hydraulic Ram.—Water flowing under the action of gravity tends to the condition in which it would be in equilibrium, and in which, therefore, all parts of the free surface would be at the same level. This is the meaning of the statement that "water seeks its own level." Usually it is only by the means of work done by some force other than gravity that water can be raised to a higher level. In the hydraulic ram a small fraction of the water in a stream is raised to a high level by a self-acting mechanism which does not need any external power.

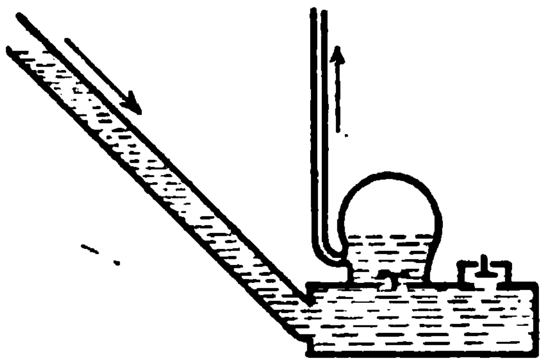


Fig. 108.—Principle of hydraulic ram.

When a stream of water in a pipe is suddenly stopped, for example when a faucet is turned off, the momentum of the water, which may be very large, is stopped in a very short time and therefore the force exerted by the water on the pipes may be very much larger than that which the water exerts after it has come to rest. In the hydraulic ram this momentary intense pressure is used to drive water into an air-chamber such as is used in a force-pump.

Momentary interruptions of the current are caused by the opening and closing of a valve which works automatically in a vertical direction. The weight of the valve is such that, when it is closed and the water is at rest, the pressure of the water on the lower surface of the valve is not sufficient to keep it closed; hence it opens and allows the stream to start. The stream when in motion carries the valve with it, again closing it and arresting the motion.

Some of the potential energy of the head of water is transformed into kinetic energy of the flowing stream, and this is partly changed into potential energy of the compressed air, which again is changed into potential energy of the water at the top of the delivery tube. Only a small part of the water is finally raised to a higher level than its original one, and its gain of potential energy is compensated by the loss of potential energy of the remainder.

Molecular Properties of Liquids

206. Molecular Forces.—Between the particles of a solid or of a liquid there are attractions that keep the body together, unless these forces are overcome by external forces. To show directly the existence of these forces between the particles of a liquid is very difficult, since a liquid so readily changes its shape. It has, however, been found possible to fill a glass tube with water at a high temperature and then seal the tube; the water, on cooling, continued to fill the tube, without contracting, until it exerted a tensile force of over seventy pounds per square inch upon the walls of the tube. The water would, in such an experiment, stand a much higher stress, if it were possible to free it perfectly from absorbed gases. It is this attraction between the particles of a liquid that has to be overcome when a liquid is evaporated; and, from the heat required for evaporation, it can be calculated that the attractions between the particles are very powerful and produce a very great internal pressure across any imaginary plane in the liquid.

From the above it might be thought that a body immersed in a liquid would feel the effect of this great internal pressure. That such is not the case is due to the fact that the molecular forces of attraction are sensible only when the distances between the particles are exceedingly small. (§160) Now the thinnest solid that it is possible to insert in a liquid separates the particles so far that the attractions between them are negligible, and thus the pressure on an immersed solid is merely that due to the causes, gravitation and pressure on the boundary, considered earlier.

The distance to which the force of attraction is sensible is called the *range of molecular forces.* Any particle of a liquid is attracted by all particles that lie within this range and these are contained within a sphere. This sphere, whose radius is the range of molecular forces, may be called the *sphere of influence* of a particle.

207. Surface Tension.—The molecular forces of which we have been speaking produce certain remarkable effects at the surface of a liquid. *The surface of a liquid tends to contract to the smallest area admissible.* Thus a drop of water falling through the air

becomes spherical, since the sphere is the figure of least surface for a given volume. The same is true of a drop of liquid lead falling in a shot tower; the drop solidifies during the fall and is found to be spherical when the fall is sufficient to allow it to become perfectly solid while in the air. A mixture of alcohol and water can be prepared of the same density as an oil, and a large drop of the mixture floating totally immersed in the oil is spherical. When the end of a stick of sealing-wax or of a glass rod is melted in a flame, it tends to the spherical form. A beautiful illustration of the tendency of a liquid surface to contract consists in forming a film from a soap-bubble solution on a ring of wire, to which a loop of silk has been loosely attached so that the loop floats in the film; when the film is broken inside the loop the latter becomes circular. In shrinking to the form of least area the film pulls the loop into the form of greatest area for a given periphery, and this is a circle.

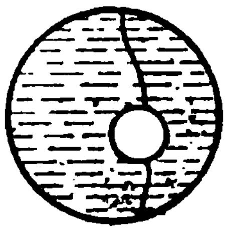


FIG. 109.—Loop of silk on surface of film.

208. Explanation of Surface Tension.—Consider the condition of a particle at A in the body of a liquid, and that of a particle at B, at less than the range of molecular forces from the surface. The particle at A is equally attracted on all sides by the particles around it, but the particle at B is more attracted inward than outward, since a sphere with center at B and the range of molecular forces as radius lies partly outside of the liquid. To take a particle from A to B, work must be done against this inward attraction.

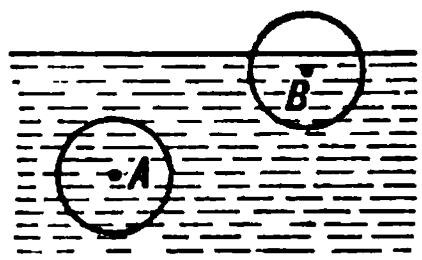


FIG. 110.

Now, when the surface of a liquid is increased, for example when a soap film is stretched, more particles are drawn into the surface; hence some work is done by the stretching force and therefore an opposing force is overcome. But the stretching force required is parallel to the surface; hence the liquid exerts an opposing or contractile force parallel to the plane of the surface, and this force is what we call the surface tension. Thus we explain the existence of a tension in the surface of a liquid by showing that it is in accordance with the principle of work. At present our knowledge of the state of the particles near the surface is too imperfect to

enable us to describe their condition more precisely and to show how the state of tension along the surface is produced.

If a line be imagined drawn along the surface of a liquid, the part of the surface on one side of the line pulls on the part on the other side, and if the length of the line be supposed one cm. the pull in dynes is taken as the magnitude of the surface tension, T, of the liquid. *The measure of surface tension is the force of contraction across a line of unit length in the surface.*

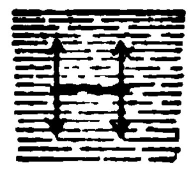


Fig. 111.—Surface tension is the force across unit length.

209. Methods of Measuring Surface Tension.—Surface tension manifests itself in many ways and, as almost any of its effects may be made the basis of a method of measuring it, the methods that have been employed are numerous. When the liquid can be formed into a thin sheet, as in the case of a soap solution, a direct method of measuring it may be used; a film may be formed on a wire frame of which one side is movable; if the force required to hold this side at rest against the surface tension is F, and the length of the movable side is l, the tension in each surface of the film is $F/2l$.

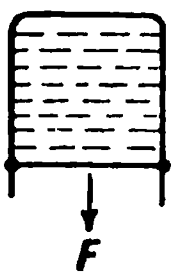


Fig. 112.—Stretching a film.

To draw a horizontal wire up through the surface of a liquid the tension of the surface must be overcome, and from the force required the surface tension may be calculated.

The movement of minute waves or ripples on the surface of a liquid is due chiefly to the surface tension of the liquid, and from the wave-lengths of the ripples and their velocities we can find the magnitude of the surface tension.

The rise of liquid in a capillary tube depends, as we shall see later, on the surface tension of the liquid, and this affords another method of measurement.

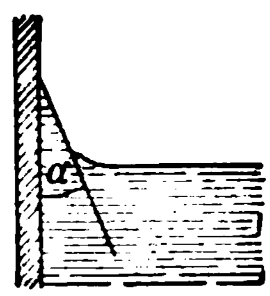


Fig. 113.—Contact of water and glass.

210. Contact of Liquid and Solid.—The general free surface of a liquid is horizontal; but, where the liquid is in contact with a solid, the surface is usually curved, the direction and amount of the curvature being different for different liquids and different solids. Water in contact with a vertical surface of glass is curved upward, and mercury in

the same circumstances is curved downward. These, for a reason stated later, are called *capillary* phenomena.

The contact angle of the wedge-shaped part of the liquid between the free surface of the liquid and the surface of the solid is called the *angle of capillarity*. The size of the angle in any case depends on the purity of the liquid and the cleanness of the solid surface. Thus for very pure water in contact with clean glass the angle is 0°; but with slight contamination, even such as is caused by exposure to air, the angle may become as large as 25° or more. For perfectly pure mercury and glass the angle is about 148°, but slight contamination reduces it to 140° or less; for turpentine it is 17°, for petroleum 26° and so on.

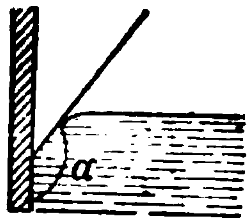


FIG. 114.—Contact of mercury and glass.

211. Level of Liquids in Capillary Tubes.—When a glass tube of very fine bore (or so-called capillary tube), open at both ends, is placed vertically with its lower end in a vessel of liquid, the surface of the liquid in the tube is usually higher or lower than the general level of the surface in the vessel. When the liquid is water or alcohol the surface is elevated in the tube; when the liquid is mercury the surface is depressed. For a given liquid the amount of elevation or depression is greater the smaller the bore of the tube, being, in fact, inversely as the diameter of the bore. For tubes of other materials than glass similar effects, depending in amount on the material of the tube, are observed.

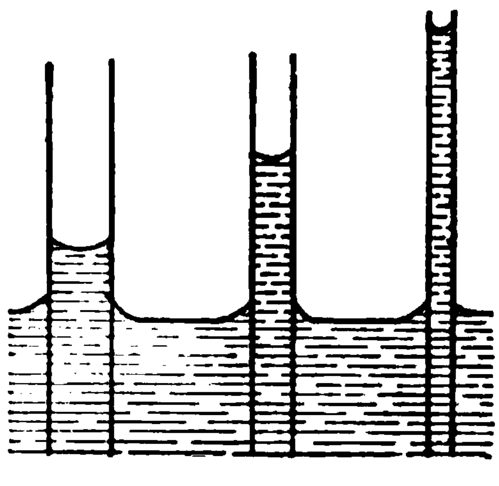


FIG. 115.—Water in capillary tubes.

There are similar elevations and depressions between two glass plates standing close together in a liquid. These elevations and depressions and the curvature of a liquid surface in contact with a solid are usually grouped under the general title of *Capillarity*.

Assuming the existence of the invariable angle of capillarity at which a liquid meets a solid, we can give a simple explanation of capillary elevations and depressions.

Consider the case when the liquid is elevated. The liquid in the tube meets the tube in a circle of radius r equal to the radius of the bore, and at every point of the circle the angle of contact

is the angle of capillarity α. Thus the surface tension of the liquid pulls on the tube in the direction PQ inclined at α to the length of the tube; and the tube therefore reacts with an equal pull in the direction QP. The amount of the pull per unit length of the circumference of the circle of contact is T, and the component of this, parallel to the length of the tube, is $T \cos \alpha$. For the whole circumference of the circle of contact the sum of these components is $2\pi r T \cos \alpha$. This is an upward force on the liquid in the tube, and it draws the liquid upward until the weight of the liquid elevated above the ordinary surface equals the supporting force. If the mean elevation is h, the volume of the supported column is $\pi r^2 h$ and its weight $\pi r^2 h \rho g$ in dynes. Hence

FIG. 116.

$$\pi r^2 h \rho g = 2\pi r T \cos \alpha$$

$$\therefore h = \frac{2T \cos \alpha}{g \rho r}$$

Thus the elevation is directly as the surface tension and inversely as the radius of the tube. By measuring the elevation and the radius and finding α by some other method, the value of T for any liquid may be obtained.

212. Elevation between Plates.—The above method of proof may also be extended to the case of a liquid between parallel plates (Fig. 116). In this case the surface of the liquid meets the surfaces of the plates in straight lines. Let the distance between the plates be d. Consider the equilibrium of the liquid contained between the plates and two vertical planes perpendicular to the plates and at unit distance apart. The pull of the surface tension at the top is $2T \cos \alpha$ and the weight of the liquid supported is $dh\rho g$. Hence

$$h = \frac{2T \cos \alpha}{g \rho d}$$

Thus the elevation is the same for two parallel plates as for a tube, if the distance between the plates equals the radius of the tube.

213. Pressure Caused by a Curved Surface under Tension.—Since the liquid in a capillary tube is elevated above or depressed below the ordinary level, the pressure beneath the curved surface must be less or greater than the pressure at the general surface.

When the effect is a depression (mercury in glass), the depressed surface is curved downward and the tension in the surface produces a pressure, just as the tension in a rubber sheet stretched over a ball produces pressure on the ball. When the effect is an elevation, the stretch on the upward curved surface tends to draw the liquid in the surface layer away from the liquid below and so produces a state of tension or diminution of pressure beneath the surface. From the amount of the elevation or depression we can calculate the change of pressure thus caused. In the case of an elevation to a height h the pressure must be less than the pressure at the ordinary level, which is atmospheric pressure, by $g\rho h$ or, (§211), $2T\cos\alpha/r$. Here r is the radius of the tube. If we denote the radius of the spherical surface by R, $R\cos\alpha=r$. Hence the pressure beneath the concave surface is less than that of the atmosphere above by $2T/R$. The same applies to the pressure produced on the concave side of a depressed surface. This difference of pressure on the two sides is due entirely to the tension and the curvature of the surface.

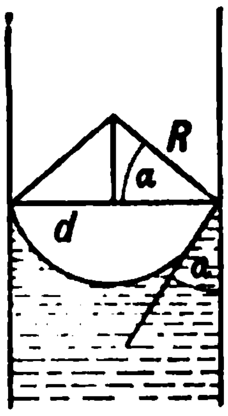


FIG. 117.

In the case of a spherical soap-bubble there are two surface tensions to be considered, one on the inner side of the film and the other on the outer side. Hence the excess pressure inside the bubble, due to the tension and curvature of the film, is $4T/R$.

A cylindrical surface in a state of tension also produces pressure on the concave side. This is deduced, as above, from the elevation or depression of a liquid of surface tension T between two parallel plates at a distance d apart. If R is the radius of the cylindrical surface of the liquid $R\cos\alpha=\frac{1}{2}d$. Hence (§211) $p=T/R$, and this is therefore the pressure on the concave side due to the tension T in a cylindrical surface of radius R. In the case of a cylindrical soap-bubble of radius R the tension in each surface produces pressure T/R. Hence the pressure inside is greater than that outside, by $2T/R$.

214. Other Effects of Surface Tension.—When the angle of capillarity of a liquid in contact with a solid is small, the liquid, in its attempt to establish this small angle, spreads out on the surface of the solid; that is, the liquid is one that wets the solid. Thus a drop of water let fall on clean glass spreads out, the angle of capillarity being small. A drop of mercury on a glass plate has no tendency to spread but gathers into a ball.

FIG. 118.—Water between glass plates.

A film of water between two glass plates makes it difficult to draw the

plates apart by a force normal to their surfaces. The liquid tends to spread over both plates and becomes concave outwards, so that the pressure within it is less than the atmospheric pressure which acts on the outside of the plates, and this produces an apparent attraction between the plates.

When an attempt is made to blow out a glass tube containing numerous detached drops a surprising resistance is experienced. Each drop becomes concave on the side of high pressure and the total resistance is the sum of the pressures exerted by these concave surfaces.

Small bodies, such as straws and sticks, floating on the surface of a liquid usually attract and gather into groups. Let us represent two such bodies by small vertical plates. If the liquid wets both it rises between them, and the pressure in the elevated portion is less than the atmospheric pressure on the outer side of the plates. Hence the plates are pushed together. If the liquid does not wet either plate it is depressed between them; the pressure above the depressed part is atmospheric, while the pressure in the liquid on the outer sides of the plates is greater than atmospheric and the plates are pushed together. If the liquid wets one plate

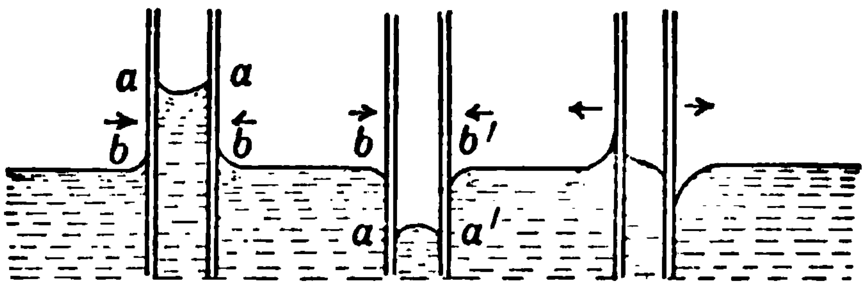


Fig. 119.—Capillary attractions and repulsions.

but not the other there is a part of each plate on which the pressure on the inside is greater than that on the outside; hence an apparent repulsion results. (Balls of paraffine wax some of which are lamp-blacked, floating on water, will illustrate all three cases.)

Any dissolved substance or impurity changes the surface tension of water. This explains the irregular motions of small particles of camphor dropped on clean water. At some points the camphor dissolves more rapidly than at other points, and near the former the surface tension of the water is weakened so that the pull on the opposite side, where the tension is greater, prevails and causes irregular motion.

215. Diffusion of Liquids.—The gradual mixture of two liquids which come into contact is called *diffusion*. It takes place on a large scale where fresh water from a river flows out into the ocean. It may be illustrated on a small scale by pouring a solution of a colored salt into a tall vessel and then cautiously covering the colored solution with a layer of water. The particles of each liquid are in motion and begin to make their way across the interface and, after a long time, the whole vessel is filled with a mixture of the same constitution throughout. Stirring has the

effect of increasing the area of contact of the liquids and so promotes diffusion. But some substances such as oil and water will not diffuse into one another, or "mix" probably because of their very different internal attractions.

Let us denote the two diffusing liquids by A and B, and let us suppose that initially A occupies the lower half of a tall jar and B the upper half. The concentration of either of the liquids at any point is its mass per unit volume at that point (*i.e.*, its density at the point if the other liquid be imagined absent without the first being disturbed). The liquid A diffuses vertically upward, that is, from places of high concentration to places of low concentration. The *gradient of concentration* in any direction is the rate at which the concentration falls off in that direction; if the rate of fall per unit of distance is unity, the gradient of concentration is unity. The general law of diffusion is that *the rate of diffusion for each liquid is proportional to the gradient of concentration* of that liquid. The *coefficient of diffusion*, or the *diffusivity* of the liquid, is the mass in grams that crosses unit area in a day when the gradient of concentration is unity. This constant can be found from observations of the density at various points along the direction of diffusion, made by means of beads of different densities floating in the liquid, and in various other ways. The following table contains the coefficients of diffusion of various substances into water at the temperature (Centigrade) stated.

Hydrochloric acid	1.74 at 5°
Common salt	0.76 at 5°
Common salt	0.91 at 10°
Sugar	0.31 at 9°
Albumen	0.06 at 13°
Caramel	0.05 at 10°

From the above it will be seen that liquids vary widely in diffusivities. Substances of high diffusivity are called *crystalloids* and those of low diffusivity are called *colloids*. The former group includes mineral acids, salts and substances generally that form crystals (whence the name), while the latter includes gums, albumens, starch, and glue (the name being derived from the Greek for glue). Crystalloids dissolved in water produce many marked changes in its properties; colloids in water form jellies,

which seem to consist of a semi-solid framework holding the liquid in its meshes. Colloids have large and complex molecules and it is, perhaps, to this fact and to the consequent slower motions of the molecules that their small diffusivities are due. They are comparatively tasteless, as they do not diffuse and reach the nerve terminals. Their low rates of diffusion also render them indigestible. Through a layer of a colloidal jelly crystalloids will diffuse almost as rapidly as through water, but colloids not at all.

216. Diffusion through Membranes. Osmosis.—Through certain membranes which have no visible pores, many liquids will diffuse readily. Thus through a partition of rubber between water and alcohol the alcohol will pass rapidly, while the passage of the water is barred. If animal membranes are wet by water, it readily passes through. A method of separating crystalloids and colloids, called *dialysis*, depends on the different rates at which these substances pass through such a membrane as parchment paper. The diffusion of substances through such septa is called *osmosis*.

Some membranes allow one constituent of a mixed liquid or solution to pass, while barring the other constituent; such membranes are called *semi-permeable*. One such is ferrocyanide of copper, formed in the pores of a porous partition by the reaction between ferrocyanide of potassium on one side and copper sulphate on the other. When such a membrane separates water and the aqueous solution of any one of various salts, the salt does not pass, but the water passes in both directions, though more rapidly toward the solution than in the opposite direction. If the solution be in a tube the lower end of which, closed by a plug of the membrane, is dipped in water, the level in the tube will rise until (provided the membrane does not break) the column is of such a height that its pressure prevents further flow. This pressure is called the *osmotic pressure* of the solution. Its magnitude, for very weak solutions, is proportional to the concentration, that is, to the number of molecules of the dissolved salt per unit volume. For a large number of salts the pressure is the same for solutions that contain the same number of molecules of the

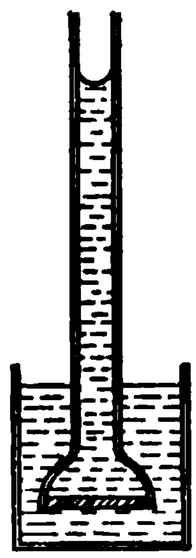


FIG. 120.—Osmotic pressure.

salt in unit volume. For various other salts the osmotic pressure for a given number of molecules per unit volume is two (or some whole number of) times greater than for the first group; this is possibly due to the molecules being resolved into atoms in the solution, the atoms acting independently. But the full explanation of osmosis and osmotic pressure is a matter of much dispute. One remarkable fact may be noted, namely, that the osmotic pressure for a given number of molecules (or of dissociated atoms) in an aqueous solution is equal to the pressure that these molecules (or atoms) would produce if freely flying as gaseous particles in the space occupied by the solution. It is also noteworthy that the osmotic pressure increases in the same way and at the same rate with rise of temperature as the pressure of a gas does.

Osmosis plays an important part in many processes that take place in the bodies of animals and plants.

PROPERTIES OF GASES

217. A gas has already been defined as a fluid which has no definite volume of its own independent of the containing vessel, but expands so as to occupy any vessel in which it is contained. Gases have the same properties as liquids in all respects which depend on the fact that the shear modulus of a fluid is zero. The pressure at a point in a gas is the same in all directions (§184). The pressure of a gas on a surface is normal to the surface (§183). Pressure applied to any part of the boundary is equally transmitted in all directions (Pascal's Principle §189). A body immersed in a gas is buoyed up with a force equal to the weight of the gas displaced (Archimedes' Principle §191). The pressure in a gas increases with its depth at a rate expressed by $g\rho h$, as in the case of liquids (§185). Gases also show the property of internal friction or viscosity, and the definition of the coefficient of viscosity of a gas is the same as that of a liquid. Some of these properties are of special importance in the case of a gas and call for separate treatment.

218. Pressure of the Atmosphere.—A very important example of the pressure of a gas is the pressure exerted by the earth's atmosphere. The atmosphere, consisting chiefly of oxygen and

nitrogen, is held to the earth by the gravitational attraction between it and the earth. The total pressure on the surface of the earth is the total attraction between the earth and the atmosphere, that is, the weight of the atmosphere. The pressure on any horizontal area of the earth's surface is the weight of all the air vertically above that area. At the top of a mountain the pressure is less than at sea level, since less of the atmosphere is above.

Galileo discovered that air had weight by weighing a glass globe containing air and then re-weighing it when he had forced more air into it. His friend and pupil Torricelli found (in 1643) that, when a tube 33 inches long filled with mercury and closed at one end was inverted in a dish of mercury, the mercury stood at a height of about 30 inches in the tube, thus leaving a vacuum above. This is known as *Torricelli's Experiment.* He thus disproved the previous view that "Nature abhors a vacuum," and was led to infer that the pressure of the atmosphere on any area equals that of a column of mercury about 30 inches high and of a cross-section equal to the area. On hearing of Torricelli's experiment, Pascal reasoned that the pressure should be less and the column of mercury in Torricelli's tube lower at the top of a mountain and he wrote to a relative, who lived near the Puy de Dome in Auvergne, to make the test. The result confirmed his conjecture.

FIG. 121 — Torricelli's experiment.

219. The Mercurial Barometer.—Torricelli's tube was the first and simplest barometer or pressure-gauge for measurement of the pressure of the atmosphere. The most accurate mercurial barometer of the present day is a Torricellian tube with a scale and vernier for accurate measurement of the height of the mercury column, and a device by which the mercury in the cistern may be readily brought to a definite height. In Fortin's *cistern barometer* (Fig. 122) the cistern, C, has a leather bottom, S, the center of which rests on a screw, V. By turning the screw the level of the mercury in the cistern can be raised or lowered so that when the barometer is read the level of the mercury in the cistern shall always be the same, namely, zero of the scale on which the height of the barometer is read. Without such an adjustment.

the level of the mercury in the cistern would fall or rise as the height of the mercury in the tube, T, rose or fell. That the level of the mercury in the cistern may be observed, the upper part of the cistern is of glass and a small ivory stud, O, projecting downward from the top of the cistern, is adjusted by the maker, so that its end is on a level with the zero of the scale. The image of the stud in the surface of the mercury is observed and when, as the level of the mercury is raised by the screw, the end of the stud and the end of its image just meet, the surface of the mercury is at the zero of the scale. In filling such a barometer care must be taken that no air remains in the mercury, and, for this purpose, after the tube has been filled it is inverted and the mercury boiled so that the air is expelled. The mercury in the cistern becomes somewhat tarnished in course of time and the image of the stud ceases to be distinct.

FIG. 122.—Cistern of Fortin's barometer.

A simpler form of barometer is Bunsen's *siphon barometer*. In this there is no cistern, but the lower end of the tube is turned vertically upward. The difference of level in the open and in the closed end is the barometric height. Thus readings of both ends of the mercury column are necessary. Scales are etched on both branches; the one on the longer arm reads upward and that on the shorter arm reads downward. The two scales are usually laid off with the same position for the zero, so that the sum of the two readings is the height of the barometer.

FIG. 123.—Bunsen's siphon barometer.

Another form of barometer is the *Aneroid* (Greek anēros = dry) barometer in which no liquid is used. It consists of a metallic box exhausted of air, with a thin metallic cover. Changes in atmospheric pressure cause slight changes of curvature in the cover, and by means of a multiplying system of levers these changes are

transmitted to a pointer, which moves around a circular scale that is graduated in cms. or inches so as to correspond to the readings of the mercurial barometer. This form of barometer is more convenient for travellers, but it has the disadvantage that its index must frequently be reset by comparison with the mercury barometer.

220. Uses of the Barometer.—A knowledge of barometric pressure is of great importance in weather forecasting. The governments of the United States and other civilized nations maintain a large number of stations where records of the barometer are kept. From simultaneous readings over a wide area the direction in which storms (or areas of low pressure) will move can be predicted. Such predictions lead annually to the saving of thousands of lives, and of much valuable property in shipping.

Since the atmospheric pressure is less at higher levels, it is possible to ascertain the height of a mountain by observing the atmospheric pressure at the top and at the bottom. Near sea-level the height of the barometer diminishes by about 0.1 inch for every 80 feet of ascent; but as the elevation increases the rate of fall diminishes owing to the greater rarity of the air. Allowance must be made for any difference of temperature at the two stations of observation.

Fig. 124.—Boyle's tube for pressures greater than atmospheric.

221. Pressure and Volume of a Mass of Gas.—Common observation shows that added pressure on a mass of gas diminishes its volume. Thus, in pumping up a bicycle tire, a large volume of air from the atmosphere is forced by high pressure into the small volume of the tire. Conversely, diminution of pressure allows a gas to expand. Against the pressure exerted on a gas it exerts an equal and opposite pressure, so that it is immaterial whether we speak of the pressure *on* or pressure *of* a gas.

The law connecting the volume and the pressure of a gas is extremely simple, but it was not discovered until 1662, the discoverer being Robert Boyle. (Fourteen years later Mariotte rediscovered the same law.) *The volume of a gas at constant temperature varies inversely as its pressure*, or, denoting the pressure and volume by P and V respectively, $PV =$ a constant. Boyle

discovered this law by experiments conducted with a tube bent as in Fig. 124, the shorter arm being closed and containing air and mercury, while the longer was open and was filled to varying depths with mercury. If to the difference of level in the two arms the height of the mercury barometer at the time is added, the sum is proportional to the pressure on the air, while the length of the tube occupied by air is proportional to the volume of the air. Thus he discovered the truth of the law for pressures exceeding an atmosphere. For pressures below an atmosphere he used a straight tube containing, initially, air and mercury and closed at one end; the open end was then plunged into a deep vessel of mercury. By drawing the tube to different heights the volume of the air increased with diminishing pressure. Thus Boyle verified the law for pressure less than an atmosphere.

FIG. 125.—Boyle's tube for pressure less than atmospheric.

222. Deviations from Boyle's Law.—While the law stated by Boyle is accurate enough for all ordinary practical purposes, careful tests have shown that it is not perfectly accurate. The most complete tests were made by Amagat. He found that in the case of air, while the pressure is being increased from one atmosphere to about 78 atmospheres, PV steadily diminishes, until its value is 0.98 of its value at one atmosphere. Thereafter, with increasing pressure, PV increases so that at 3000 atmospheres it has a value 4.2 times its initial value. In the first stage (that is, up to 78 atmospheres) V decreases more rapidly than Boyle's law would indicate; thereafter it decreases less rapidly, so that at 3000 atmospheres its volume is 4.2 times what it would be if Boyle's law were perfectly accurate. (It may be noted that at 3000 atmospheres air has a density of 0.93, nearly equal to that of water; while the density of liquid oxygen at its critical pressure is about 0.7 and that of liquid nitrogen about 0.4.)

Other gases, show similar deviations from Boyle's law; but the pressure at which PV is a minimum is widely different for different gases, and so, too, is the magnitude of this minimum value of PV.

In his earlier experiments (1881) Amagat measured pressures by a very tall manometer in a mine shaft. Later he designed a special gauge for

very high pressures. This consisted of two opposed pistons of very different diameters in separate cylinders. The high pressure, P, was applied to the small piston, of area a, and was counterbalanced by a much smaller pressure, p, applied to the large piston, of area A. Evidently $Pa = pA$, and, p being measured by a mercury manometer, P was deduced. Very viscous liquids were used in the cylinders to diminish leakage.

Starting with the view that a gas consists of flying particles the impact of which produces the pressure observed in a gas, Van der Waals deduced the following formula which agrees very well with the results of Amagat's experiments.

$$(P + a/V^2)(V - b) = \text{a constant},$$

at constant temperature, a and b being constants that are different for different gases.

223. Modulus of Elasticity of a Gas.—The shear modulus of a gas being zero, a gas has only one modulus, namely the bulk modulus, and this is (when the gas is kept at constant temperature) simply equal to the pressure, P, of the gas. This is seen from Boyle's law. For when the pressure is P and the volume V, let an additional small pressure p be applied and let the volume be thereby reduced by the small quantity v, then by Boyle's Law

$$(P+p)(V-v) = PV$$

or if we neglect the product of the small quantities p and v

$$Vp = Pv$$

Now the bulk modulus is the increase of pressure p divided by the proportional decrease of volume v/V, and from the last equation this is equal to P.

224. Buoyancy of a Gas.—A body such as a balloon, lighter than the volume of air which it displaces, will ascend in the air when released. The force giving it an acceleration upward equals the difference of its weight and the weight of the air which it displaces. If it rises to such a height that its mean density equals the density of the rarefied atmosphere, it will not ascend unless lightened by casting some of its load overboard. A large man displaces about $\frac{1}{4}$ lb. of air. When a body is weighed in air with weights that are supposed correct if used in a vacuum, the true weight of the body will not be obtained unless correction be made for the effect of the buoyancy of the air.

225. Manometers.—A manometer is an apparatus for measuring the pressure of a fluid. In the simplest form the pressure to be measured is balanced against the pressure of a column of liquid in a tube. This is called the *open tube manometer* or siphon gauge. The pressure is found from the difference of level of the liquid in the two arms and the density, ρ, of the liquid. In absolute units of force $P = g\rho h +$ atmospheric pressure, while in the weight of unit mass as unit of force $P = \rho h +$ atmospheric pressure.

In another manometer the pressure to be measured is balanced against that of a gas (usually air) in a uniform *closed tube*. By Boyle's Law the pressure in the gas is inversely as the volume, that is, inversely as the length of the air column. The pressure in the gas plus that indicated by the difference of level of the liquid is the pressure to be measured.

FIG. 126.—Open tube manometer.

In *Bourdon's Pressure Gauge* a hollow tube of metal having an elliptical cross-section is bent into an arc of over 180°. One end of the tube is closed. When the fluid of which the pressure is to be measured is admitted to the open end, the curved tube will become less curved under increased pressure and more curved under decreased pressure. An index moving over a scale is attached to the free end. The action depends on the fact that the pressure tends to increase the interior volume of the tube; and, since a circular cross-section allows of more volume than an elliptical one for a given periphery, the section will under increased pressure tend to the circular form and the change of form of the cross-section causes the change of shape of the tube.

FIG. 127.—Closed tube manometer.

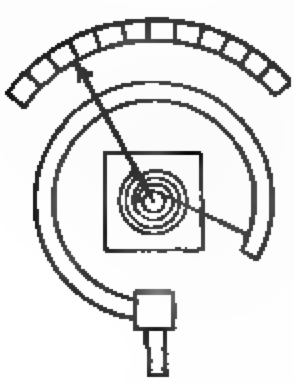


FIG. 128.—Bourdon's pressure gauge.

226. Viscosity of Gases.—The viscosities of gases are small compared with those of liquids. Thus the viscosity of air is about $\frac{1}{60}$ of that of water. While the viscosity of air is small, it is sufficient to

retard greatly the fall of small particles of dust and small drops of water such as constitute a cloud. In a cloud (where the air may be one thousand times less dense than water) a drop of water one thousandth of an inch in diameter falls about 0.8 inch per second, while a drop one ten-thousandth of an inch in diameter falls about one hundred times more slowly, or about 0.5 inch in a minute. For large drops such as constitute rain the viscosity of air offers practically no resistance; the resistance which prevents such drops attaining enormous velocities is the inertia of the air.

The viscosity of a gas increases when its temperature rises, which is the opposite of the case with liquids. The viscosity of a gas at constant temperature does not change appreciably when its density is altered by change of pressure.

227. The Kinetic Theory of Gases.—The view that a gas consists of a myriad of particles in incessant motion may be regarded as firmly established. The evidence for this belief is that we can from it deduce nearly all the properties of a gas, and the agreement between these deductions and the observed facts could hardly be a mere accidental coincidence. As we do not yet know the details of the structure of the particles of which a gas consists, there are some properties of a gas which we cannot yet deduce from this theory. A definite contradiction between the numerous known properties of a gas and the deductions made from the theory would be fatal to the latter; no such contradiction has ever been found.

As an illustration of the way in which the theory accounts for the properties of gases we shall show that it explains Boyle's Law.

Before doing so we must state the theory more in detail. The following, while an incomplete statement, will be sufficient for our purposes. (*a*) A single gas consists of particles all of the same size moving in random directions; (*b*) when the particles impinge on one another and on the walls of the vessel, they rebound like smooth spheres with a coefficient of restitution of unity; (*c*) unless a gas is greatly condensed, the particles are so far apart compared with their dimensions that the forces they exert on one another may be neglected except at impact. It will be noticed that we do not assume that the velocities of all the particles are the same, and, in fact, there is good ground for believing that the velocities differ considerably.

For simplicity, consider a gas contained in a rectangular vessel the edges of which are a, b and c in length, and let A_1 and A_2, each of area bc, be

perpendicular to the edges of length a. Let us first fix our attention on some particular particle which has a velocity V in some direction. V may be resolved into three components u, v, w, in the directions of the edges respectively, u being in the direction of the a sides. Suppose the particle to impinge on the side A_1. The force that it will exert on that side at impact will depend on its mass and on u, not at all on v and w. If it impinged without rebounding it would give momentum equal to its mass, m, multiplied by u, or mu. But it rebounds with a velocity the component of which perpendicular to A_1 is u in the opposite direction; hence the momentum it gives to A_1 is $2mu$. Let us now suppose, for the present, that it reaches A_2 without impinging on any other particle; for this it will require a/u seconds. At A_2 it will rebound with a velocity the component of which perpendicular to A_2 is u, and will, supposing it to encounter no other particle, reach A_1 in time $2a/u$ when it will again rebound. Hence in every second it will impinge $u/2a$ times on A_1, and in every second it will give to A_1 momentum $2mu \cdot u/2a$, or mu^2/a. The total force exerted on A_1, that is the momentum imparted to A_1 per second, is the sum of mu^2/a for all the particles, and, to find the pressure p on A, we must divide this sum by the area of A_1, namely bc. Hence

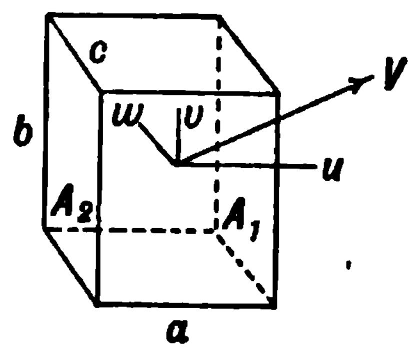


FIG. 129.

$$p = \frac{m}{abc}(u_1^2 + u_2^2 + \ldots)$$

Let us now denote the total number of particles in the vessel by N, and the number per unit volume by n. Since abc is the total volume of the vessel, $nabc = N$. Hence

$$p = mn\frac{(u_1^2 + u_2^2 + \ldots)}{N}$$

The product mn is the mass of all the particles in unit volume, that is, the density ρ; and $\dfrac{u_1^2 + u_2^2 + \ldots}{N}$ is the average value of u^2 for all the N particles in the vessel. Denoting this by $\overline{u^2}$ we see that $p = \rho\overline{u^2}$. For any one particle

$$V^2 = u^2 + v^2 + w^2,$$

and, since the particles are moving wholly at random, the average values of u^2, v^2 and w^2 are all equal and the value of each is therefore $\frac{1}{3}$ of the average value of V^2 which we may denote by $\overline{V^2}$. Hence

$$p = \tfrac{1}{3}\rho\overline{V^2}.$$

If v be the volume of a mass M of gas, since $\rho = M/v$,

$$pv = \tfrac{1}{3}M\overline{V^2}$$

The total kinetic energy of translation of the gas is the sum of the kinetic energies of translation of all the particles and is evidently equal to $\frac{1}{2}MV^2$ or $\frac{3}{2}pv$. Now there is good reason to believe that if the temperature of a gas is constant, this kinetic energy is constant. Hence the product of the pressure and volume of a gas at constant temperature is constant, and this is Boyle's Law.

In the above we have neglected the fact that a particle may, during its passages between A_1 and A_2, impinge on other particles. If such an impact take place between two particles moving along a line perpendicular to A_1 and A_2, the particles will exactly exchange their velocities (§174), since they are of the same mass; and the second particle will, therefore, have in the a-direction a component equal to that of the first particle before impact. Thus the second particle will take the place of the first in the process described above. When the immense number of particles and the random nature of their motions are considered, it is seen that the effect is the same as if all the impacts were in the directions of u, v, and w.

The deviations from Boyle's Law are due to the (very small) forces between particles when they are not in contact. These we have neglected; by considering them Van der Waals arrived at his more correct law.

228. Surface Condensation and Occlusion.—When a gas is in contact with a solid there are molecular forces drawing the particles together, and these produce more or less condensation of the gas on the surface of the solid. This makes it impossible to remove the last traces of a gas from a glass vessel by means of an air pump. It also accounts for the fact that, when a figure is traced on a sheet of glass by a stick, the figure will appear when the glass is breathed on. The breath condenses less readily on the part of the glass that has been freed from condensed gas by the scraping of the stick.

A porous solid is readily permeated by a gas and condensation on the surfaces of the pores takes place. This is called *occlusion*. Very porous wood-charcoal will absorb nine volumes of oxygen, thirty-five volumes of carbonic acid and ninety volumes of ammonia per volume of the charcoal, and cocoanut-charcoal will absorb still more. This is why charcoal is so useful as a deodorizer. Platinum in the porous form called platinum sponge will absorb 250 times its own volume of oxygen. Palladium will absorb more than one thousand volumes of hydrogen. Its own volume is thereby increased by about one-tenth. The hydrogen is therefore reduced to one thousandth of its original volume; to

produce such a condensation by pressure alone would require a pressure of several tons per square inch.

229. Diffusion of Gases.—Gases, because of their greater mobility, diffuse much more rapidly than liquids. When two vessels containing different gases are connected by a wide tube, diffusion proceeds with great rapidity, and in a short time each gas is found distributed in both vessels as if the other gas were not present. If one of the gases be a colored gas, such as chlorine, the process of diffusion can be observed. As regards the final result each gas acts to the other as a vacuum, but in the process of diffusion each gas retards the other. Gravity also plays some part in the process though not in the final result. Thus if the gases be hydrogen and carbon dioxide, the final mixture is attained more rapidly when the carbonic acid is in the higher vessel.

In the process of diffusion of two gases into each other each gas follows the same law as holds for the diffusion of two liquids, that is, each gas diffuses from places where the concentration of that gas is great to places where it is less, and the rate of diffusion is proportional to the rate of fall of concentration.

230. Efflux of Gases.—The rate of escape of a gas through a small aperture in a very thin plate may be deduced from the principle of energy. Each part of the gas as it escapes has a certain velocity and therefore a certain kinetic energy, and this must equal the work performed by the pressure in the vessel in forcing the gas out. Let P be the excess of the pressure in the vessel over the external pressure. During the escape of a small volume V of the gas the pressure P does the same amount of work as if it had pushed out a piston in a cylinder. Hence (§195) the work done is PV. If the density is ρ the mass of the volume V of the gas is $V\rho$, and if its velocity is v its kinetic energy is $\frac{1}{2} V\rho v^2$. Equating the work done to the kinetic energy which it produces, we get

$$v=\sqrt{\frac{2P}{\rho}}$$

Thus the rate of escape is directly as the square root of the pressure and inversely as the square root of the density.

Bunsen's method of comparing the densities of gases consists in comparing their rates of escape through the same aperture under the same pressure.

In establishing the above formula we have supposed that no work is done against internal friction such as there would be if the escape were through a tube. The wall of the vessel was supposed very thin so that the diameter of the opening might be larger than the thickness of the wall. Yet even in this case there is some slight viscous friction. This friction is different for different gases; hence the above simple formula does not give the ratio of the densities very accurately. When a mixed gas escapes by effusion the composition of the escpaing gas is not altered as it escapes.

When a gas escapes through a porous partition in which the pores are very small, such a fine unglazed pottery-ware, the circumstances are different from those of the above cases. The pores are comparable in size with the molecules of the gas and, as might be expected, the rates of escape of different gases are so different that the constituents of a mixed gas escape at different rates. This affords a method of partially separating the constituents of a mixed gas, and, as the process may be repeated several times, the separation may be made nearly complete. By this process it has also been possible to show that the molecules of a single gas are all of the same size, since no separation can be produced by the above method.

231. Passage of a Gas through Rubber.—Some gases also escape through membranes such as rubber and wet parchment, in which there are no pores in the ordinary sense. The gas is dissolved by the membrane on one side and given up on the other side, so that the passage through the membrane is a diffusion from parts of the membrane where the concentration is greater to parts where it is less. The same is true of the passage of a gas through a film of liquid. In a somewhat similar way hydrogen will pass through red-hot platinum and iron.

232. Pumps for Liquids.—The oldest form of pump, or *suction pump*, consists of a piston moving in a cylinder or barrel which is connected with the well by a pipe. In the pipe, or at the top of the pipe, there is a valve, called the inlet valve, which can open towards the cylinder, but not in the opposite direction; and in the piston there is a valve, called the outlet valve, which can open outward but not inward toward the cylinder. When the piston is first raised the air in the cylinder expands and its pressure diminishes. The outlet valve closes owing to the excess of

FIG. 130.—Suction pump.

pressure on the outside, and, for the same reason, the inlet valve opens and air from the suction pipe enters the cylinder. Thus the air in the suction-pipe is rarefied and the greater atmospheric pressure on the water in the well forces water some distance up the suction-pipe. After some strokes the water enters the cylinder and flows out by the outlet valve.

Since it is the pressure of the atmosphere that raises the water in a suction-pump, water cannot be raised by this means higher in the pipe than atmospheric pressure will raise water in a vacuum; and, since the density of water is to that of mercury as 1 to 13.6, it follows that the maximum theoretical height is 13.6 times the height of the mercury in a barometer or about 34 feet. The practical limit of suction-pumps is considerably less than this, owing to the presence of air in water and to the difficulty of making the contact between piston and pump air-tight. When water is to be raised higher a *force-pump* is used (Fig. 131). This differs from the suction-pump in the fact that the outlet valve is not in the piston but in a side tube connected to the cylinder near the inlet valve. During each downward stroke of the piston water is forced up this side tube, and the height that may be reached will depend on the force that can be applied to the piston and the maximum pressure that the pump will stand without breakage of some part.

The outflow from the delivery tube of a force-pump as described above would be intermittent; but it may be rendered more nearly continuous by means of an "air chamber," in which air, being put under pressure by the water forced in, exerts continuous pressure on the outflowing water.

Fig. 131.—Force pump

233. The Siphon.—The siphon is a bent tube for removing liquid from a vessel. The tube is filled with liquid and is then inverted, and one end A is immersed in the liquid, while the other end C is kept closed. When C is opened liquid flows through the

tube and out through C, so long as C is below the level, D, of the surface of the liquid.

To explain the action of the siphon let us consider the pressure on the liquid at C before the end C is opened. If the difference of level of D and C is h, the pressure on the liquid at C is greater than atmospheric pressure by $g\rho h$. Hence, when C is opened, the excess of pressure inside causes a flow, and the flow continues so long as C is below the level of D and A remains immersed. A siphon will not act if the highest point B of the tube is at a greater height above the level of D than the height to which atmospheric pressure will force the liquid in an exhausted tube.

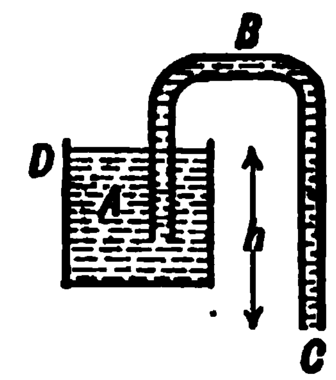


FIG. 132.—Siphon.

234. Air-pumps. The first pump for removing air from a vessel was invented by Otto von Guericke (in 1650). It was essentially a suction pump like that used for water, the only difference being the closer fit of piston required to prevent leakage in the case of a gas. The degree of exhaustion that can be attained by such a pump is low. The flap-valve, at the end of the suction-tube, will not act automatically when the pressure in the receiver has become very small. For this reason a conical plug, carried by a rod that passed with some friction through the piston, was substituted. Another difficulty is caused by the fact that the piston cannot be made to fit the lower end of the cylinder with perfect accuracy, so as to expel all the air drawn from the receiver into the cylinder. The latter defect has been remedied in the Geryk pump (Fig. 133) which has a layer of oil at the bottom of the cylinder; oil above the piston also prevents leakage at the piston valve.

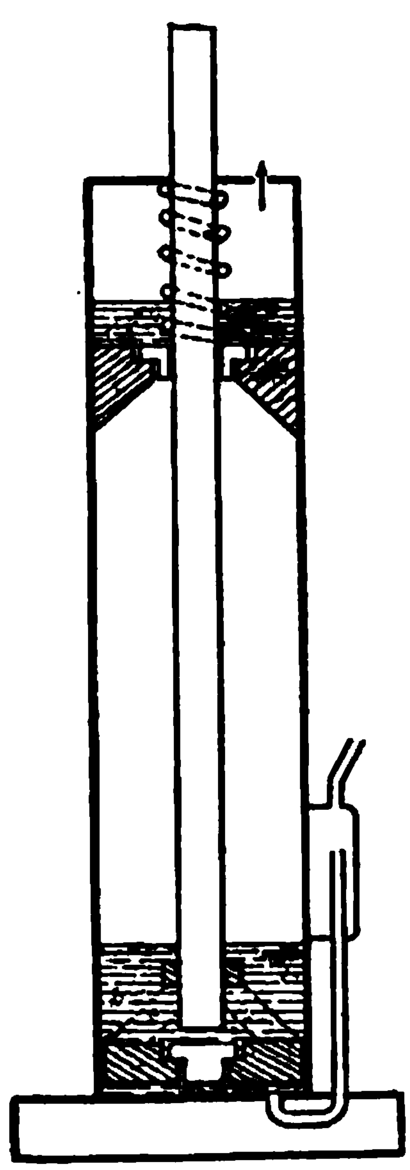


FIG. 133.—Geryk pump.

235. High Vacuum Pumps.—Many pumps have been devised for removing gases from vessels and obtaining very high vacua. In nearly all of them mercury has been used. In the older forms the level of mercury in a large bulb connected to the receiver was alternately lowered

and raised so that the gas was drawn from the receiver into the bulb and then ejected through a side tube, or the mercury fell in drops through a narrow tube and exerted suction on a side tube connected to the receiver. These forms of mercury pump are rapidly going out of use and we shall describe only two of the most recent and efficient mercury pumps.

Langmuir's "mercury vapor pump"[1] makes use of the principle of the aspirator (§ 193) in a novel form. A current or "blast" of mercury vapor passes upward from the heated flask, *A*, through the tubes *B* and *C* (Fig. 134) into the condenser *D*. An annular space, *E*, surrounding *B* is connected through *F* with the vessel to be exhausted. *C* is enlarged into a bulb, *H*, just above the upper end of *B* and *H* is surrounded by a condenser, *J*, through which water flows. The mercury condensing in *D* and *H* returns to *A* through the tubes *L* and *M*. The gas from *F* passes freely up through *E* and, meeting the blast of mercury vapor at *P*, is blown outward and upward along the walls of the condenser *H* and forced into the main stream of mercury vapor passing up through *C* into the condenser *D*. A less efficient pump connected to *N* maintains a vacuum of about 0.3 mm. (400 bars) and removes the gas. Langmuir's pump will exhaust a vessel of 11 liters capacity from atmospheric pressure to a vacuum of 0.00001 mm. (0.015 bars) in 80 seconds. Because of its remarkable simplicity and rapidity of action it marks a great advance in methods of obtaining high vacua.

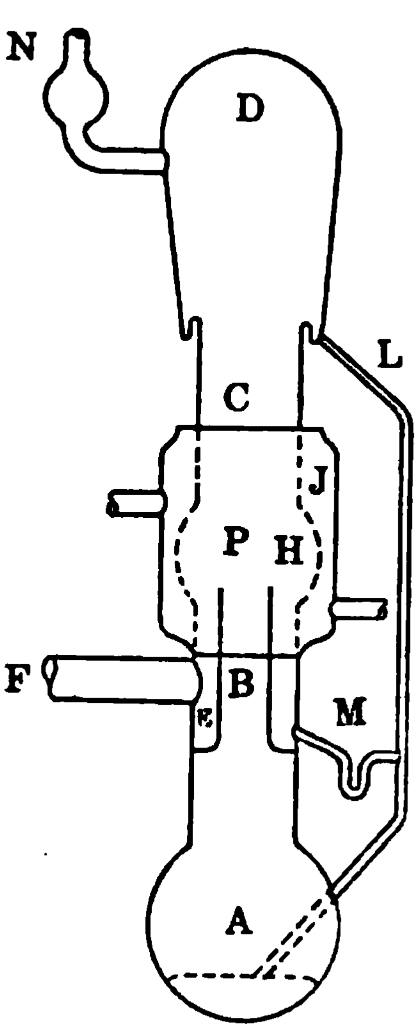


FIG. 134.—Langmuirs Mercury vapor pump.

The principle of Gaede's mercury pump is indicated (without details) in Fig. 135 and Fig. 136. An iron cylinder, *g*, with a glass face, *g'*, is more than half filled with mercury, the surface of which is at *q*. Inside of *g* there is a porcelain drum, *t*, rotating about an axis, *a*, which passes air-tight through *g*. This drum is divided into two chambers, w_1 and w_2, which com-

[1]The substance of this description has been supplied, with great courtesy, by the inventor of the pump, Dr. Irving Langmuir. A more complete account will appear in an early number of the Physical review.

municate with g by long channels between the division-walls of t. Each chamber has an opening, f, by which the part of the chamber above the mercury is connected, through the tube R, with the receiver to be exhausted. As the drum is rotated counter-clockwise, the chamber w_1 is gradually emptied of mercury and filled with air drawn in through R. As the rotation continues, f_1 is immersed, and the air in w_1 is driven into g. The action of w_2 is similar. Since either f_1 or f_2 is always out of the

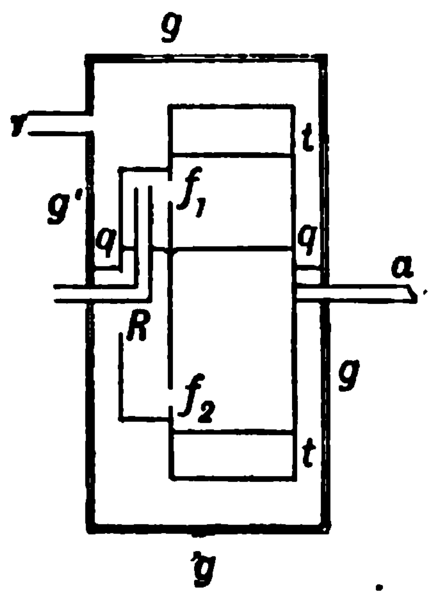


Fig. 135.

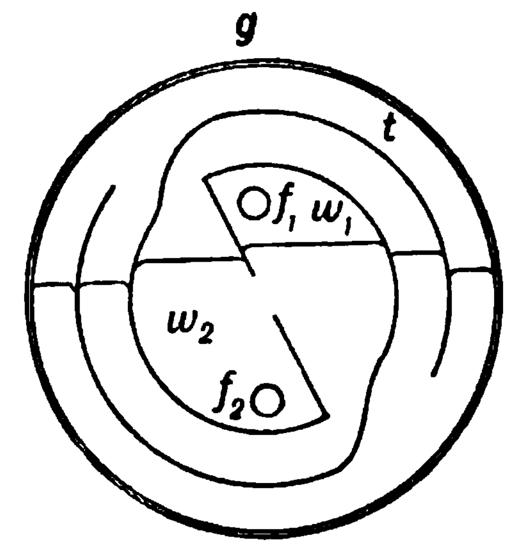


Fig. 136.

mercury, the suction through R is continuous. The air in g is removed by another pump (which may be much less efficient) connected to r. Gaede's pump will produce a vacuum of about 0.00004 mm.

References

CREW'S *Principles of Mechanics* contains a brief and very clear account of the subject stated in elementary Vector and Calculus language.

POYNTING & THOMSON'S *Properties of Matter* is especially valuable for information on gravitation, elasticity and properties of fluids.

The above-mentioned books will be found useful for somewhat advanced systematic study.

MACH'S *Principles of Mechanics* is a very interesting and elementary account of the historical development of the subject.

COX'S *Mechanics* is an elementary book with notes on the historical development.

WORTHINGTON'S *Dynamics of Rotation* is an elementary book with numerous suggestive experiments.

PERRY'S *Spinning Tops* is a popular account of the principles of the gyroscope.

LOVE'S *Theoretical Mechanics* is a very careful account of the logical relations of the parts of the subject.

MAXWELL'S *Matter and Motion*, while elementary and very brief, is a masterpiece by a great physicist.

TAIT'S *Properties of Matter* contains an elementary treatment of gravitation, elasticity and properties of fluids.

LODGE's *Pioneers of Science* consists of popular lectures on Galileo, Newton, etc.

POYNTING's *Mean Density of the Earth* describes all the methods used.

Encyclopedia Britannica, articles on "Weights and Measures," "Mechanics," "Elasticity," etc.

DANIELL's *Principles of Physics* is a large compendium.

Problems

Velocity and Acceleration.

1. A train acquires 5 minutes after starting a velocity of 40 km. per hour. Assuming constant acceleration, what is the distance passed over during the 5th minute? *Ans.* 0.6 km.
2. A train having a speed of 70 km. per hour is brought to rest by brakes in a distance of 600 m. What is the acceleration (assumed constant)? *Ans.* -1.13 km./min^2.
3. What is the final speed of a body which, moving with uniform acceleration travels 72 meters in 2 minutes if:
   (a) the initial speed = 0?
   (b) the initial speed = 15 cm. per sec.?
   Ans. 120 cm./sec.; 105 cm./sec.
4. A body is projected at an angle of 30° with the horizon with a velocity of 30 m. per sec. When and where will it again meet this horizontal plane? How far will it ascend? *Ans.* 3.06 s.; 79.5 m.; 11.4 m.
5. A body slides down an inclined plane and in the 3d sec. describes 110 cm. What is the inclination? *Ans.* 2° 35′.
6. What initial vertical velocity must a ball have in order to fall back to its starting point in 10 sec.? *Ans.* 4900 cm /sec.
7. At what angle with the shore must a boat be directed in order to reach a point on the other shore directly opposite, if the speed of the boat be 4 miles per hour and that of the stream be 2 miles per hour? *Ans.* 60°.
8. A point goes over a circular path 10 cm. in diameter 4 times a second, at a uniform speed. To what acceleration is it subject? *Ans.* 3158 cm./sec^2.
9. A ball rises to a height of 50 ft. and travels 200 ft. horizontally. Find the direction and magnitude of the velocity with which it is thrown *Ans.* $\theta = 45°$; $v = 80.2$ ft./sec
10. Show that the time of descent (without friction) down all chords of a vertical circle is the same.
11. What velocity must a boy give a sling of 80 cm. radius in order that the stone shall not fall out of the sling? *Ans.* 280 cm./sec.

Force and Mass.

12. What force will a man who weighs 70 kg. exert upon the floor of an elevator descending with an acceleration of 100 cm. per sec. per sec.? If ascending with the same acceleration? *Ans.* 62.8 kg. wt.; 77.1 kg. wt.
13. A force of 1000 dynes acts upon a mass of 1 kg. for 1 min. Find the velocity acquired and the space passed over in this time. *Ans.* 60 cm./sec.; 1800 cm.

14. A shot weighing 10 lbs. is shot from a gun weighing 3 tons with an initial velocity of 1200 feet per sec. What is the initial velocity of the recoil? *Ans.* 2 ft./sec.

15. Three forces, 5, 12, 15 are in equilibrium. Find the angles between them. *Ans.* 62° 11′; 134° 58′; 162° 51′.

16. Bodies of mass 10 kg. and 8 kg. are connected by a string over a pulley. How far does each move from rest in the first two seconds? *Ans.* 218 cm.

17. Twelve bullets are divided between two scale pans connected by a cord passing over a pulley. What distribution will produce the greatest tension on the support of the pulley?

18. A cord passes over two fixed pulleys and through a third pulley suspended between them. A mass of 10 g. is attached to one end of the cord, a mass of 5 g. to the other end, and the suspended pulley and the attached weight weigh 2 g. The cords being all vertical, what are the accelerations of the three masses? *Ans.* 809 cm./sec^2.; 639; 724.

19. A baseball whose mass is 300 g. when moving with a velocity of 20 m. per sec. is squarely struck by a bat and then has a velocity of 30 m. per sec. in the other direction. Calculate the impulse and average force if the contact last .02 sec. *Ans.* 15×10^5 g.-cm./sec.; 7.5×10^7 dynes.

20. With how much energy must a bullet weighing 20 g. be shot horizontally from a gun 4 m. above a level plane, in order to strike the ground 300 m. away from the gun? *Ans.* 1.11×10^{10} ergs.

Work and Energy.

21. A projectile traveling at the rate of 700 ft. per sec. penetrates to the depth of 2 in. Find the velocity necessary to penetrate 3 in. *Ans.* 857 ft./sec.

22. A hammer weighing 6 kg, and moving with a velocity of 100 cm. per sec. drives a nail into a plank 1 cm. What resistance does it overcome (supposed uniform)? *Ans.* 3×10^7 dynes.

23. A man can bicycle 12 miles an hour on a smooth road; downward force of each foot in turn = 20 lbs., length of stroke = 1 ft., bicycle is advanced 17 ft. for each revolution of the cranks. At what H. P. does he work? *Ans.* .075 H. P.

A man, weight 180 lbs., runs up 26 steps, each 7 in. high, in 4 sec. At what H. P. does he work? *Ans.* 1.2 H. P.

25. A sprinter who weighs 161 lbs. runs 40 yds. in $4\frac{4}{5}$ sec., 60 yds. in $6\frac{3}{5}$ sec., 100 yds. in 10 sec. What is (a) his velocity from 40 to 60 yds. and from 60 to 100 yds., (b) his kinetic energy at the end of 40 yards? (c) Calculate the rate of working in H. P. required to produce this kinetic energy. (d) In what other ways does he expend energy?
Ans. (a) $33\frac{1}{3}$ ft./sec. (b) 2777 ft. lbs. (c) 1.09 H. P.

26. Find the number of watts in one horse-power. *Ans.* 746.

27. A sprinter does 100 yards on the horizontal in 10.5 sec., and the same distance up hill with a rise of 32 ft. in 17.5 sec. Assuming that his rate of working is the same throughout, calculate the added work done in the additional 7.0 seconds up hill and the rate of working that this implies, the man's weight being 160 lbs. *Ans.* 1.33 H. P

28. 100 cu. ft. of water pass over a dam 10 ft. high in 1 min. What horse-power could be derived from this if all were utilized? *Ans.* 1.9 H. P.

29. A 30-gram rifle bullet is fired into a suspended block of wood weighing 15 kilos. If the block is suspended by a string of length 2 meters and is moved through an angle of 20°, calculate the velocity of the bullet. Notice that the impact of the bullet on the block does not change the total *momentum* of both (§ 46) and during the subsequent swing of the pendulum its total *energy* remains constant. *Ans.* 770 m./sec.

30. If a locomotive driving-wheel 1.5 m. in diameter makes 250 revolutions per minute, what is the mean linear speed of a point on the periphery? Of the point when it is highest? When it is lowest? *Ans.* 19.6 m./sec.; 39.2 m./sec.; 0.

Rotation.

31. The armature of a motor revolving at the rate of 1800 revolutions per minute comes to rest in 20 seconds after the current is shut off. Calculate its average angular acceleration and the number of revolutions. *Ans.* -9.42 rad./sec.^{2}; 300 rev.

32. Find in radians per second the angular velocity of the earth about its axis and deduce the component of this angular velocity about a diameter through a point in latitude 40°. (Principle of Foucault's pendulum).

33 A circle has a diameter of 16 cm. A smaller circle tangent to it and 12 cm. in diameter is cut out of it. Where is the center of gravity of the remainder? *Ans.* 10.6 cm. from common tangent.

Center of Mass.

34. Two cylinders of equal length ($=20$ in.), and having diameters of 12 and 6 in., are joined so that their axes coincide. Where is the center of gravity? *Ans.* 6 in. from junction.

35. Find the center of gravity of a table 4 ft. × 3 ft. × 1 in., with legs at the corners 2 ft. × 2 in. × 2 in. *Ans.* 0.233 ft. from top.

36. The mass of the moon is $\frac{1}{80}$ of that of the earth and the average distance between their centers is 240,000 miles. Calculate the position of the center of mass of the two. *Ans.* 2963 m. from center of earth.

37. At the corners of an equilateral triangle ABC masses of 1, 2 and 3 lbs. respectively are placed. Find the distance of their center of mass from BC assuming each side of the triangle to be 1 ft. in length. *Ans.* 0.144 ft.

38. A bar 6 ft. long and pivoted at the middle has a weight of 24 lbs. hung at one extremity. What is the moment of the weight (a) when the bar is horizontal, (b) when it makes an angle of 30° below, and (c) of 60° above with the horizontal position? *Ans.* 72; 62.3; 36 lbs. wt. ft.

Moments.

39. If it is wished to upset a tall column by a rope of given length pulled from the ground, where should it be applied, if the length of the rope is, (1) equal to, (2) twice, the height of the column?

40. Find the moment of inertia of a sphere ($m=20$, $r=2$) about an axis tangent to its surface. *Ans.* 112.

41. Find the moment of inertia of three circular disks, all three touching each other in the same plane, about a perpendicular axis passing through

the center of one of them. The mass of each is 100 g. and the radius of each is 6 cm. *Ans.* 34,200 gm. cm.^{2}

42. Two masses, 100 kg. and 200 kg., respectively, are connected by a rigid rod 1 m. long. The system is thrown so that the center of gravity has a velocity of 20 m. per second and the system turns 10 times per second about this center. Find the kinetic energy of the system. *Ans.* 192×10^{10} ergs.

43. What energy has a grindstone $1\frac{1}{2}$ m. in diameter, weighing 1000 kg. and rotating once every 2 sec.? *Ans.* 13.9×10^9 ergs.

44. A solid iron cylinder, 100 cm. diameter, rolls down a plane 6 m. long inclined at 30°. What linear velocity does it acquire? *Ans.* 627 cm./sec.

45. A block of stone weighs 2.5 tons and is in the form of a cube of 1 yard side. It rests on level ground. What is the least force which applied to the block will cause it to revolve about a horizontal edge? *Ans.* 1768 lbs. wt.

46. Parallel forces of 1, 2 and 3 units respectively act at the corners A, B, C of an equilateral triangle of 1 ft. side. Find the distance of the resultant from BC. *Ans.* 0.144 ft.

Parallel forces.

47. Parallel forces of 10 and 6, but in opposite directions, are applied to a bar at distances of 8 and 3 from one end. What is the magnitude of the resultant and where does it act? *Ans.* 4; 15.5.

48. Two equal parallel forces, each 50 dynes, act in opposite directions at the ends of a bar 10 cm. long. The bar makes an angle of 45° with the direction of the force. What is the moment of the couple? *Ans.* 353.5 dynes-cm.

49. A man and a boy carry a weight of 20 kg. between them by means of a uniform pole 2 m. long, weighing 5 kg. Where must the weight be placed so that the man may carry twice as much of the whole weight as the boy? *Ans.* 0.416 m. from middle.

50. A rod, the mass of which is 1 kg., hangs from a hinge on a vertical wall and rests on a smooth floor. Calculate the force on the floor and the force on the hinge. *Ans.* 500 g.; 500 g.

Equilibrium.

51. A uniform ladder 30 feet long and of 50 lbs. weight rests with the upper end against a smooth vertical wall, and the lower end is prevented from slipping by a peg. If the inclination of the ladder to the horizontal is 30°, find the force on the wall and at the peg. *Ans.* 43.3 lbs. wt.; 66.1 lbs. wt.

52. A barn door is 10 ft. long and 5 ft. wide and weighs 200 lbs. The hinges are 1 ft. from the ends and the weight is carried entirely by the upper hinge. Find the direction and magnitude of the resultant force on the upper hinge. *Ans.* 209 lbs. wt.; 17° 21′ to vertical.

53. One end of a certain rod is clamped. If the other end is pulled 1 cm. from its natural position and then released, it starts with an acceleration of 10 cm. per sec. per sec. What is the period of its vibration? *Ans.* 1.98 sec.

Periodic motions.

54. The balance-wheel of a watch makes 5 complete vibrations in 2 sec. With what angular acceleration will it start when turned 30° from its position of equilibrium and released? *Ans.* 129.34 rad./sec².

55. A hoop of 25 cm. radius hangs on a peg. Prove that its period of vibration is equal to that of a simple pendulum whose length is equal to the diameter of the hoop.

56. A clock gains 3 min. a day. Find the error in the length of the pendulum, regarding it as a simple pendulum. ($g = 980$). *Ans.* 0.414 cm.

57. A pendulum which is a seconds pendulum where $g = 980$, vibrates but 59.95 times a minute on top of a mountain. What is the acceleration of gravity at this point? *Ans.* 978.37.

58. A rod 2 m. long is freely suspended at one end. Calculate its period of vibration. *Ans.* 2.31 sec.

59. A seconds pendulum is drawn aside and released and at the same moment a ball is allowed to fall. The ball and the bob collide as the pendulum passes through the vertical. Calculate the height of fall of the ball. *Ans.* 122.5 cm.

60. The coefficient of friction for two surfaces = 0.14. A pull of 20 kg. weight will overcome what pressure between them? *Ans.* 143 kg.

Friction.

61. What force applied parallel to a plane inclined at 20° will push up a block weighing 100 kg., the coefficient of friction between the two being 0.24; (*a*) the block moving uniformly; (*b*) the block having an acceleration of 100 cm. per sec. per sec.?

Ans. (*a*) 56.7 kg. wt.; (*b*) 66.9 kg. wt.

62. What is the coefficient of friction between a body and a horizontal plane if the body loses a velocity of 100 ft. per sec. and comes to rest in moving 200 ft. over the plane? *Ans.* 0.776.

63. A toboggan slides 100 yards down a track inclined at 20° to the horizontal in 11 seconds. Calculate the coefficient of friction. *Ans.* 0.20.

64. A small block rests on a horizontal revolving platform at a distance of 40 cm. from the axis of revolution. If the coefficient of friction is .30 at what angular velocity of the platform will the block just begin to slip? *Ans.* 2.71 rad./sec.

Machines.

65. A man raises a stone 1 in. with a lever of the first class 10 ft. long weighing 50 lbs., the fulcrum being 1 ft. from the point of application to the stone. If he exerts a force of 100 lbs. wt. what force is applied to the stone and what work does he do? *Ans.* 1,100 lbs. wt.; 75 ft. lbs.

66. A boy who exerts a push of 50 lbs. wt. wishes to roll a barrel weighing 200 lbs. into a wagon $2\frac{1}{2}$ ft. high. Assuming that he pushes in a line through the center of the barrel parallel to the plank, how long a plank will he need and how much work will he do?

Ans. 10 ft.; 500 ft. lbs.

67. A body weighs 12 lbs. on one side of a false balance and 12.5 lbs. on the other side. What is the ratio of the arms of the balance?

Ans. 1.021.

68. A man weighing 150 lbs. sits on a platform suspended from a movable

pulley and raises himself by a rope passing over a fixed pulley. Supposing the cords are parallel, what force does he exert?

Ans. 50 lbs. wt.

69. A wheel whose radius is 25 cm. is fastened to one end of a screw whose pitch is 1 mm. What force can the screw exert in its nut when a force of 1 kg. wt. is applied tangentially to the wheel, friction being supposed negligible? *Ans.* 1570 kg. wt.

70. Compare the mechanical advantages of a block and tackle when the end of the cord is attached to the upper block and when it is attached to the lower.

71. How far above the surface of the earth must a body be to lose 0.1 per cent. in weight? *Ans.* 1.95 mi.

Gravitation.

72. If the moon's mass is $\frac{1}{80}$ that of the earth, and its diameter 2160 miles, that of the earth being 7900 miles, what is the acceleration of gravity on the moon's surface?

Ans. 164 cm./sec^2.

73. Find the time of revolution of the earth which would cause bodies to have no apparent weight at the equator. *Ans.* 1.41 hr.

74. A wire 300 cm. long and 1 mm. in diameter is stretched 1 mm. by a weight of 3000 g. What is Young's Modulus?

Ans. 11.2×10^{11} dynes/cm.^{2}.

Elasticity.

75. A weight is hung from the ceiling by a steel wire 2 m. long and of 1 mm. diameter joined to a copper wire 1 m. long and of 0.5 mm. diameter. Another weight sufficient to produce a total extension of 1 mm. is added. Calculate the extension of each part. *Ans.* 0.19 mm.; 0.81 mm.

76. To opposite faces of a cubical block of jelly of 20 cm. edge parallel and opposite forces of 1 kg. each are applied and produce a relative motion of 1 cm. Calculate the strain, the stress and the shear modulus.

Ans. 0.05; 2450 dynes/cm.^{2}; 49000 dynes/cm.^{2}

77. An iron bar of 400 c.c. volume falls from a ship and sinks to the bottom of an ocean 1000 m. deep. How much is its volume diminished, assuming that each 10 m. of water pressure produces a pressure equal to that of the atmosphere, which equals one million dynes per sq. cm.

Ans. 0.026 c.c.

78. A ball weighing 20 kg., moving with a velocity of 500 cm. per sec., strikes a second ball weighing 100 kg. which is at rest. If the first ball rebounds with a velocity of 100 cm. per sec., what will be the velocity of the second? *Ans.* 120 cm./sec.

79. Two bodies differing in bulk weigh the same in water; compare the weights in mercury; in vacuo.

Properties of Liquids.

80. A mass of copper suspected of being hollow weighs 523 g. in air and 447.5 g. in water. What is the volume of the cavity? *Ans.* 16.8 c.c.

81. The specific gravity of ice is 0.918, that of sea-water 1.03. What is the total volume of an iceberg of which 700 cu. yds. is exposed?

Ans. 6438 cu. yds.

82. A block of wood weighing 1 kg., whose specific gravity is 0.7, is to be loaded with lead so as to float with 0.9 of its volume immersed. What weight of lead is required (1) if the lead is on top? (2) if the lead is below? *Ans.* 286 g.; 313.5 g.

83. A hydrometer sinks to a certain mark in a liquid of sp. gr. 0.6, but it takes 120 g. to sink it to the same mark in water. What is the weight of the hydrometer? *Ans.* 180 g.

84. One of the limbs of a U-shaped glass tube contains mercury to the height of 0.175 m.; the other contains a different liquid to a height of 0.42 m., the two columns being in equilibrium. Required, the specific gravity of the second liquid with reference to mercury and to water.

85. Find the volume in cu. ft. of the smallest block of ice which, floating on fresh water, will just carry a man who weighs 150 lbs. *Ans.* 29.3 cu. ft.

86. Given a body A which weighs 7.55 g. in air, 5.17 g. in water, and 6.35 g. in another liquid B; required, the specific gravity of the body A and the liquid B. *Ans.* 3.17; 0.504.

87. A block of brass 10 cm. thick floats on mercury. How much of its volume is above the surface, and how many cm. of water must be poured above the mercury so as to reach the top of the block? (Density of mercury $=13.6$; of brass $=8.5$.) *Ans.* 0.375 of the whole; 4.05 cm.

88. Two tubes are inserted in a vessel of water on the same horizontal plane. The diameter of the one is 0.5 mm. and its length is 20 cm.; the diameter of the other is 0.25 mm. and its length is 10 cm. Compare the amounts of water flowing through the two tubes in a given time. *Ans.* 8 : 1.

89. The diameter of the small piston of an hydrostatic press is 2 in., the diameter of the large piston is 2 ft. What weight on the small piston will support two tons on the large piston? *Ans.* 27.77 lbs.

90. The pressure at the bottom of a lake is three times that at a depth of 2 m. What is the depth of the lake? (Atmospheric pressure $=76$ cm. of mercury.) *Ans.* 26.67 m.

91. A retaining wall 3 m. wide and 40 m. long is inclined at $30°$ to the horizontal. Find the total force in kg. exerted against it by the water when the water rises to the top. *Ans.* 9×10^4 kg.

92. What is the outward force exerted by the water on the sides of a circular tank 1 m. in diameter, the height of the water being 150 cm.? What is the thrust due to the water on the bottom? *Ans.* 3532 kg. wt.; 1178 kg. wt.

93. The surface tension of a soap-bubble solution is 27.45 (dynes/cm.). How much greater is the pressure inside a soap-bubble of 3 cm. radius than in the air outside? *Ans.* 36.6 dynes/cm^2.

94. How far will water be projected horizontally from an aperture 3 m. below the water level of a tank and 10 m. above the ground (neglecting air resistance)? *Ans.* 10.96 m.

95. A body whose specific gravity is 2 is weighed in air of specific gravity 0.0013 with weights of specific gravity 9. The weight in air being 100 g., what is the true weight? *Ans.* 100.050 g.

Properties of Gases.

96. If the barometer sinks 15 mm., how much is the pressure in dynes per sq. cm. decreased? *Ans.* 19992 dynes/cm^2.

97. An air bubble at the bottom of a pond 6 m. deep has a volume of 0.01 c.c. Find the volume just as it reaches the surface, the barometer standing 760 mm. *Ans.* 0.0158 c.c.

98. Owing to the presence of air the mercury column in a barometer 85 cm. long stands at 70 cm. when an accurate barometer stands at 75 cm. What pressure will this barometer indicate when an accurate barometer stands at 72 cm.? *Ans.* 67.67 cm.

99. A barometer reads 73 cm. Calculate the thrust on one side of a board 1 m. square. *Ans.* 9928 kg. wt.

100. A barometer has a cross-section of 2 sq. cm. and is so long that as the mercury stands at 76 cm., there is a vacuum space 10 cm. long. Some air is allowed to enter and the mercury falls 10 cm. What was the volume of the air before it entered? *Ans.* 5.26 cm^3.

101. How high must we ascend above the sea-level to observe a depression of 1 mm. in the height of the barometer? Density of air = 0.0013 (approx.). *Ans.* 10.4 m.

102. A glass tube 60 cm. long, closed at one end, is sunk, open end down, to the bottom of the ocean. When drawn up it is found that the water has penetrated to within 5 cm. of the top. Atmospheric pressure = 76 cm. of mercury. Calculate the depth of the ocean, assuming the density constant, and equal to 1.026. (Principle of Lord Kelvin's sounding apparatus.) *Ans.* 110.8 m.

103. In a vessel of 1 cu. meter volume are placed the following amounts of gas: (1) hydrogen, which occupies 1 cu. m. at atmospheric pressure. (2) nitrogen, which occupies 3 cu. m. at a pressure of 2 atmospheres. (3) oxygen, which occupies 2 cu. m. at a pressure of 3 atmospheres. Calculate pressure of mixture. *Ans.* 13 at.

104. The mouth of a vertical cylinder 18 in. high is closed by a piston whose area is 6 sq. in. If a weight of 100 lbs. be placed on the piston, how far will it descend, supposing the atmospheric pressure to be 14 lbs. per sq. in., the friction negligible and the temperature constant? *Ans.* 9.8 in.

105. A cylindrical diving-bell 7 ft. in height is lowered until the top of the bell is 20 ft. below the surface of the fresh water. If the barometer height at the time is 30 in., how high will the water rise in the bell? What air pressure in the bell would just keep the water out? *Ans.* 2.96 ft.; 1.82 at.

106. (a) What fraction of an atmosphere is the difference in pressure between two points in air at 0° C. and 76 cm. pressure if the difference of level is 1 cm.? (b) How large a difference of level would produce a difference of pressure of 0.01 per cent. of an atmosphere? *Ans.* 126×10^{-8}; 80 cm.

WAVE MOTION

By E. Percival Lewis, Ph. D.

Professor of Physics in the University of California

236. Characteristics of Wave Motion.—The word wave recalls the familiar phenomena observed whenever the surface of a body of water is disturbed. Large waves are usually so irregular that it would be difficult to reach any general conclusions regarding the laws of their formation or propagation. If less complex waves be observed, such as those produced by throwing a pebble into a quiet pond or by the gentle disturbance of the water or mercury in a tank, it will be seen that they are alternate ridges and hollows in the surface, which diverge in uniformly expanding circles from the center of disturbance. If small pieces of cork rest on the surface another important characteristic of wave motion may be observed. The particles rise on an approaching wave, ride forward on its crest for a short distance, then fall back into the succeeding hollow, to again move upward and forward on the next crest. They describe orbits in a vertical plane which are evidently circular or elliptical. Since these particles participate in the movement of the water on which they rest, it is plain that the water as a whole does not move continuously forward with the waves, but that each element rotates about its original undisturbed position, to which it returns when the train of waves has passed. Waves are, therefore, the progression of a *shape* or *condition*, not of matter.

237. Water waves illustrate the following fundamental characteristics of all wave motions in material media: (1) *All parts of the medium reached by the disturbance are subject to periodic displacements about their positions of equilibrium.* (2) *The disturbance is propagated at a uniform rate, each displaced particle transferring its motion to its neighbors by pressure or through some mechanical connection.* The moving elements of the medium possess kinetic energy due to their motion and potential energy

due to their displacements. This energy, originally derived from the source of disturbance, is passed on from element to element, so that there is a continuous *flow of energy* with the advancing waves.

238. Types of Waves.—The displacements in the case of water waves do not extend far beneath the surface, hence disturbances are propagated in two dimensions only, in superficial waves. There is another familiar type, resembling water waves in general shape, which may be propagated along a linear medium, such as a wire or rope. These may be called linear waves (although the disturbance extends across a finite area) because they are propagated in one direction only. Such waves may be studied by filling a long rubber tube with shot and suspending it from a tall support, holding the lower end taut in the hand. If the tube is struck a sharp blow near the lower end, a distortion resembling a wave crest will travel slowly to the upper end, where it will be immediately reflected with reversed curvature, on account of the elastic reaction at the fixed point. (Fig. 137, *a*, *b*, *c*). It will travel to the lower end and be reflected back and forth several times until its energy is exhausted by friction. This is a **solitary** wave. If the lower end is rapidly moved back and forth through a small amplitude, by properly timing the displacements a series or train of waves of opposite curvatures ("crests and hollows") will travel upward, crossing a similar train reflected downward. The combined effect of the two trains is to cause the tube to oscillate between the positions shown by the full and the dotted line in Fig. 137, *d*.

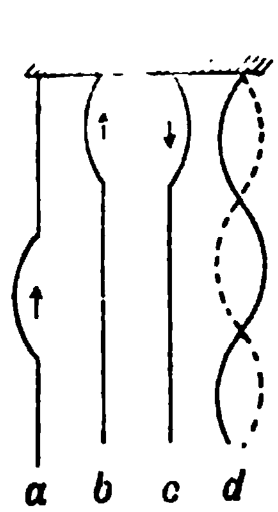


FIG. 137.

In the cases mentioned the oscillations of the medium are in part or altogether at right angles or *transverse* to the direction of propagation, so that the displacements of the boundary of the medium give rise to a definite wave shape. It is possible, however, for the vibrations to take place in the direction of propagation of the wave, as is the case with one component of the displacement in water waves. When the displacements are altogether in the direction of propagation it is evident that the wave can have no shape, as the boundaries of the medium are not displaced, but there will be periodic changes in density, arising

from the fact that different particles are at any instant in different phases of displacement, so that in one region they will be crowded together, while in another they will be separated. This may be illustrated by a row of massive spheres, connected by elastic cords or springs, as shown in Fig. 138, *a*. If the second sphere were immovable, the first alone would oscillate when pulled downward and released. If the spheres are all free to move, the transmitted impulse will set all in vibration. On account of the inertia of the spheres and the elasticity of the connections, the displacement of each sphere will lag behind that of its neighbor below, and each vibration will be in a different phase, until we come to the sphere *B*, which begins its first vibration when *A* begins its second vibration. The figure shows the resultant effect when the first sphere has completed one vibration (*b*) and one and a half vibrations (*c*) after it first moved upward through its resting point. It is evident from the figure that the *conditions* of condensation and of rarefaction are propagated with the velocity of the wave. There is no change of shape in the system, but if lines proportional to the displacements are drawn from each resting point, to the right for upward displacements, to the left for downward displacements (that is, if each displacement is rotated through 90° to the right or the left), a smooth curve drawn through the ends of these lines will have the general shape of a transverse wave (*b*, *c*). We have thus a means of graphically representing *longitudinal* waves in a way clearly coördinating them with *transverse* waves.

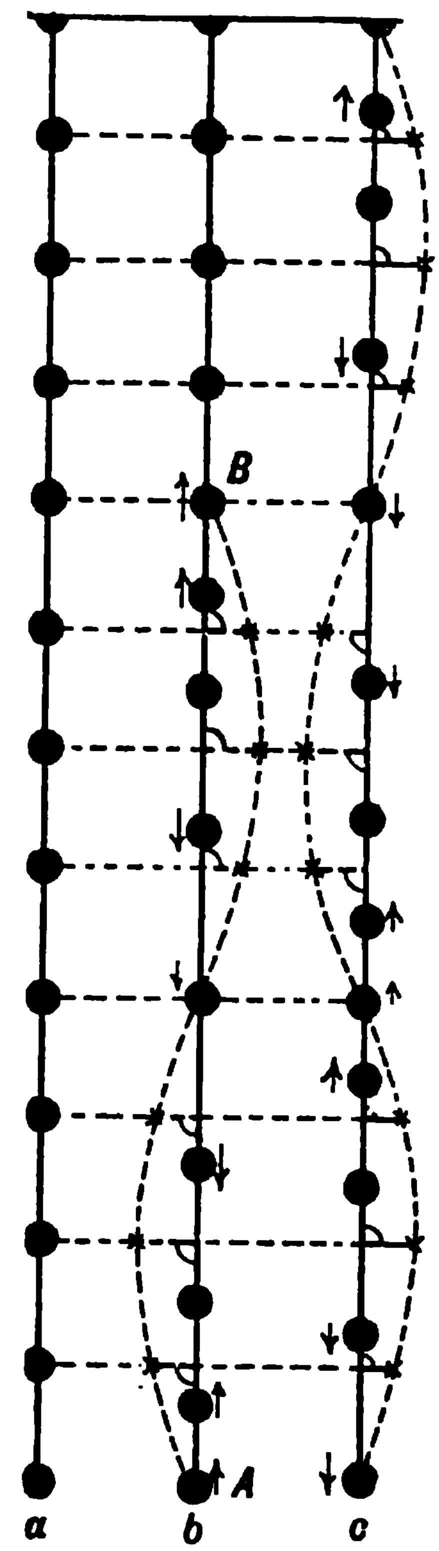


FIG. 138.

If a series of heavy bars are attached horizontally at equal intervals to a suspended wire, and if the lowest bar executes torsional vibrations, waves of angular displacement will travel up the wire. Such *torsional* waves may be represented graphically

by erecting ordinates proportional to the angle of torsion at each point on an axis representing the wire.

There are many cases where wave disturbances, such as those of sound in air, are propagated in three dimensions in a uniform medium. These disturbances will travel equal distances in all directions in equal times, hence the waves will be *spherical*, with the source as a center. A hemispherical wave of this type would be produced in a block of rubber by striking it at a point.

So far we have considered the effect of mechanical disturbances of a medium only. The idea of wave motion may, however, be extended to cases where any physical condition in a medium varies periodically at each point and is propagated with a finite velocity through the medium. A familiar example is found in the "heat waves" which travel into the earth as a result of the periodic heating and cooling of the surface. In the afternoon the surface reaches a maximum temperature. Owing to the slow conduction of the heat, this maximum travels slowly downward, all the while becoming less and less, owing to the fact that each particle passes on only a portion of the energy received by it, not nearly all, as in the case of elastic media. At night the surface reaches a minimum temperature which penetrates into the soil at the same rate as the maximum. The distribution of temperatures in the afternoon and at night are represented by the full and the dotted line in Fig. 139. The abscissa of the point A represents the average temperature. AB is the distance traveled by the heat wave in twenty-four hours. Another example of immaterial waves is found in the electrical waves traveling along conductors or in free space, due to periodic change in the electrical condition at different points. Light waves are very short electrical waves (§543).

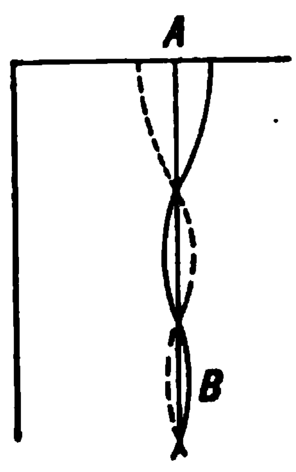


Fig. 139.

239. Vibrations in Wave Motion.—In all departments of physics, particularly in Sound, Light, and Electricity, waves play an important part, hence the study of wave motion is of fundamental importance. Since periodic displacements or changes in condition are an essential feature of wave motion, it is necessary to study such phenomena in detail. The only periodic motions which lend themselves readily to simple analysis are those of uniform motion in a circle or the projections of such motions along a line, the latter being called simple harmonic motions. (§108 *et seq.*)

As pointed out in §111, the vibrations of all elastic bodies must

be either simple harmonic motions or compounded of such motions (§248), since, for small displacements at least, the force of restitution is proportional to the displacement.

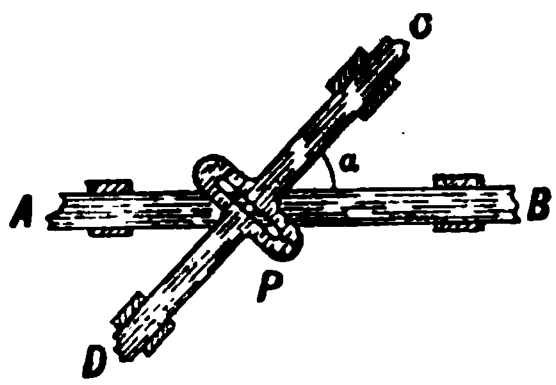


FIG. 140.

240. Resolution of Simple Harmonic Motions.—As the motion is a linear displacement, it may be resolved into two or more components like any other displacement (§25). If, for example, the piston rod AB (Fig. 140) executes simple harmonic vibrations in a horizontal line (the projected motion of the crank pin on a fly-wheel), a pin P attached to it and sliding in a slotted cross bar attached to the rod CD will cause the latter to execute a simple harmonic vibration in the direction of its length, if guides allow it to move only in that direction. If the amplitude of AB is r, the length of the crank arm, that of CD is $r \cos \alpha$.

241. Superposition of Simple Harmonic Motions.—In many cases a body may be subjected to several simultaneous simple harmonic displacements in the same or in different directions and of the same or different periods. Familiar illustrations are found in the vibrations of musical instruments (§605 *et seq.*) and whenever different sets of waves are superimposed on or cross each other. If the displacements are entirely independent, it is evident that the resultant effect may be obtained by the geometrical addition of displacements (§13). If a light pendulum is suspended from a heavy one, as shown in Fig. 141, and both set in vibration in the same plane, at a given instant the total displacement of the lower bob is $x = x_1 + x_2$; or if the pendulums vibrate at right angles, the resultant displacement is $r = \sqrt{x^2 + y^2}$. In such a case the two systems are not entirely independent, on account of their connections and inertia, and the two displacements will not remain of the simple harmonic type. If a simple pendulum be set in vibration, and later an impulse at right angles to its direction of motion be applied, it will move in a circular or elliptic orbit (conical pendulum), or in a line inclined to its original direction. In studying these effects the most useful cases to consider are those in which the periods of the components are either equal or in some simple ratio to one another.

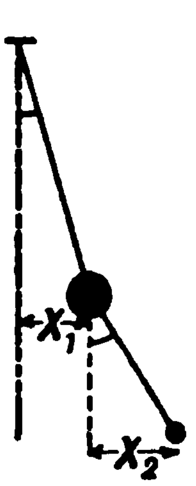


FIG. 141.

242. Composition of Two Simple Harmonic Motions of Same Period and in Same Line.—A body at O (Fig. 142) has a horizontal simple harmonic motion of period T and amplitude r_1. When the phase of this vibration is e a second simple harmonic vibration of the same period, in the same line, and of amplitude r_2 is imparted to the body. When the phase of the second disturbance becomes ωt that of the first is $\omega t+e$; e is the **phase difference.** To find the resultant displacement and phase, describe circles of reference of radii r_1 and r_2 about O. On these radii, including the angle e, complete the parallelogram, and draw the diagonal $OC=R$. Then, denoting OP by x_1, OQ by x_2, and OC by R,

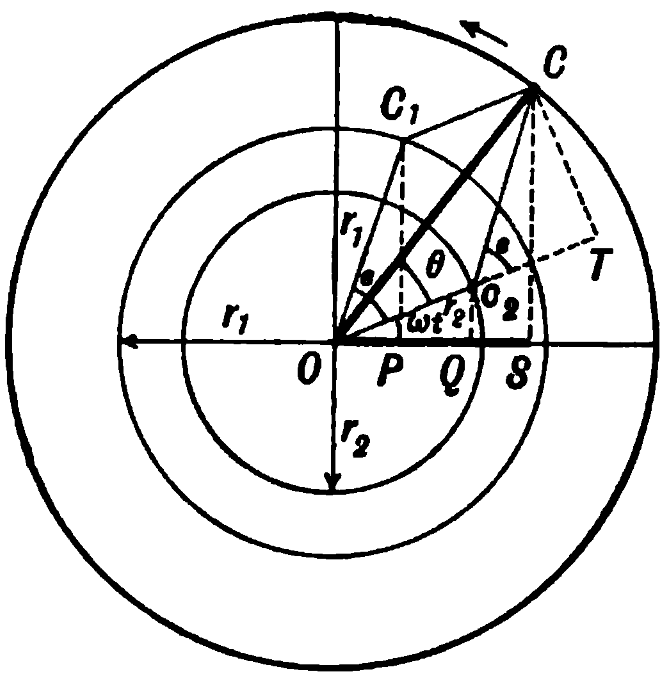


FIG. 142.

$$x_1=r_1\cos(\omega t+e) \qquad (1)$$

$$x_2=r_2\cos\omega t \qquad (2)$$

and, from the parallelogram OC_1CC_2,

$$R^2=r_1^2+r_2^2+2r_1r_2\cos e \qquad (3)$$

If x is the resultant displacement

$$x=x_1+x_2=OP+OQ=OQ+QS=OS.$$

Hence
$$x=R\cos(\omega t+\theta) \qquad (4)$$

This holds good for any value of ωt. The resultant is, therefore, *a simple harmonic motion of the same period as that of the components, of amplitude R and with a phase $(\omega t+\theta)$ intermediate between the phases of the components.* It is the projected motion of the point C in the resultant circle of reference of radius R.

The above results may also be obtained from the component simple harmonic motions (1) and (2) without use of the circles of reference. For the sum of x_1 and x_2 may evidently be written in the form:

$$x=(r_1\cos e+r_2)\cos\omega t-(r_1\sin e)\sin\omega t.$$

If we now introduce a new length R and a new angle θ such that

$$R\cos\theta=(r_1\cos e+r_2)$$
$$R\sin\theta=(r_1\sin e)$$

we can by simple trigonometry obtain (3) and (4) and also an expression for $\tan\theta$.

It is evident from (3) that R is a maximum, (r_1+r_2), when $e=0$ and R is a minimum, (r_1-r_2), when $e=180°$.

While the above refers to the addition of two simple harmonic motions of the same period, we can extend it to the case of two vibrations of different periods by supposing the phase difference, e, to change uniformly with the time. We may suppose the two motions to start at the same instant, e being then 0. At time t, the value of e will be $(\omega_1-\omega_2)t=2\pi(n_1-n_2)t$, where n_1 and n_2 are the respective frequencies. When $(n_1-n_2)t=0, 1, 2, 3,$ etc., $\cos e$ will be 1 and, from (3), R will be a maximum (r_1+r_2). When $(n_1-n_2)t=\frac{1}{2}$, $\frac{3}{2}$, etc., R will be a minimum (r_1-r_2). The interval between two successive maximum values of R is $1/(n_1-n_2)$ and the number of maxima per second is (n_1-n_2). This case is illustrated by "beats" in Sound (§600).

243. Composition of Two Simple Harmonic Motions of Same Period at Right Angles.—If the amplitudes of the respective vibrations are r_1 and r_2, construct a rectangle with sides $2r_1$ and $2r_2$, the equilibrium position of the vibrating particle being at the center. Construct two circles of diameters $2r_1$ and $2r_2$, as shown in Fig. 143. The projections on the X and Y axes respectively of points moving uniformly around these circles of reference will give the x and y components of the displacements of the body. If the former is in advance of the latter by the phase angle e, the body will be at P when the y displacement begins. Divide each circle into the same number of equal parts, beginning at C_1 and C_2, and number these in regular order. It is evident that the successive positions of the body will be at the intersections of the lines 1–1, 2–2, 3–3, etc., and a smooth curve drawn through these

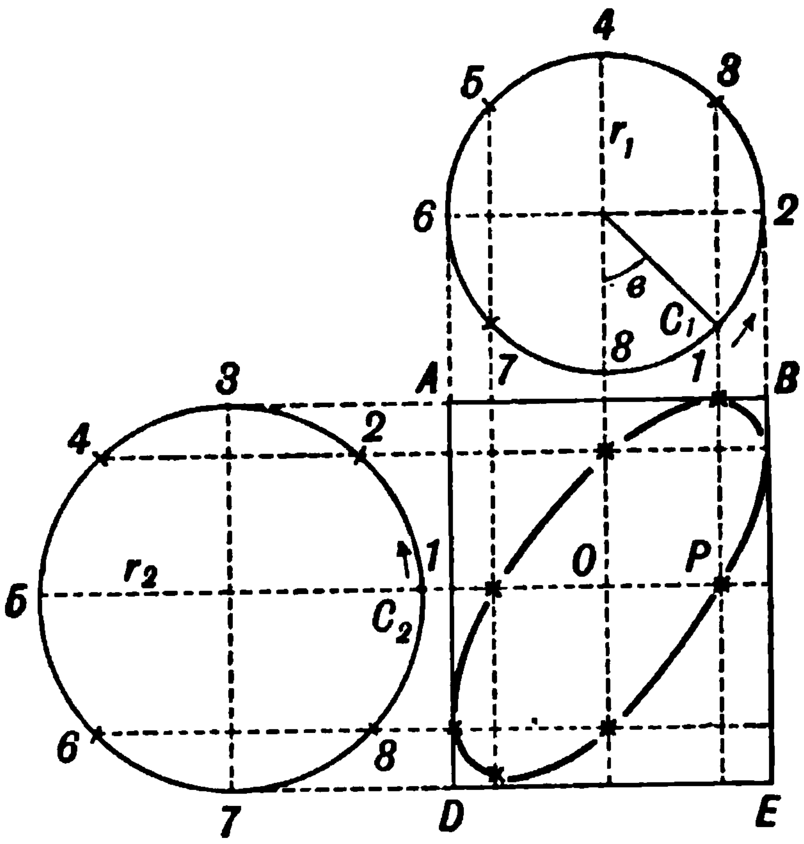


FIG. 143.

points will give the orbit of the body. In the case illustrated, where $e=45°$, this path is an ellipse inclined to the axes. If the phase difference is zero, the path is a straight line, the diagonal BD. If $e=90°$, the path is an ellipse with vertical and horizontal axes, or a circle if $r_1=r_2$. Orbits corresponding to different values of e are shown in the top row of Fig. 145.

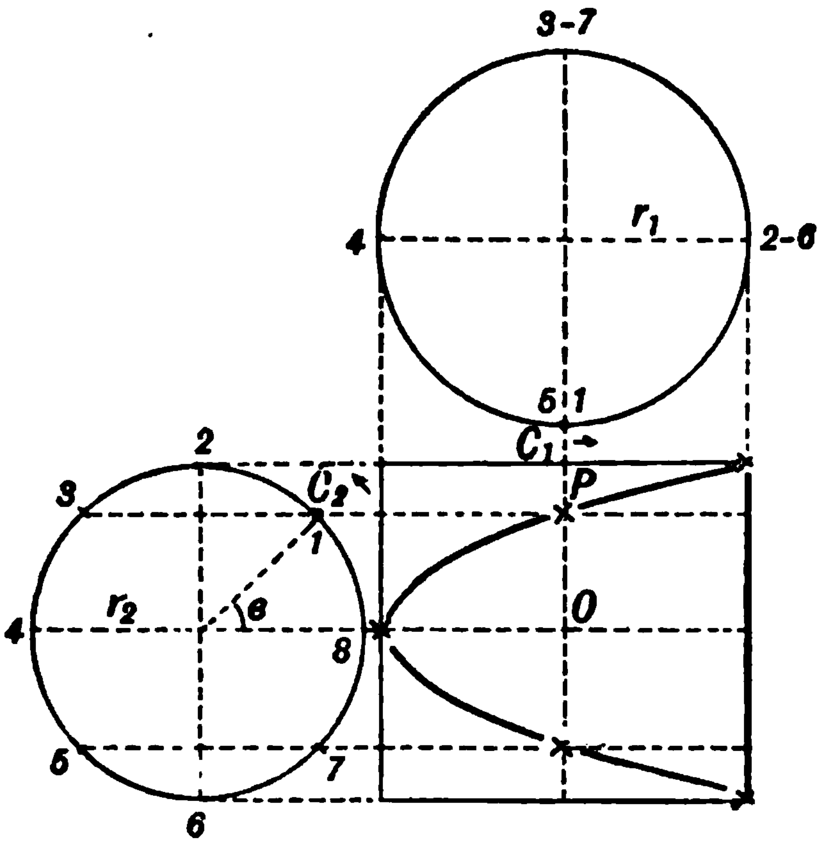


FIG. 144.

If the periods differ slightly, one vibration will gain on the other in phase, and the orbit will run through the complete cycle of forms shown in the top row of Fig. 145. If n_1, n_2 are the respective frequencies, the cycle will repeat itself whenever one component gains a whole vibration on the other, or n_1-n_2 times a second.

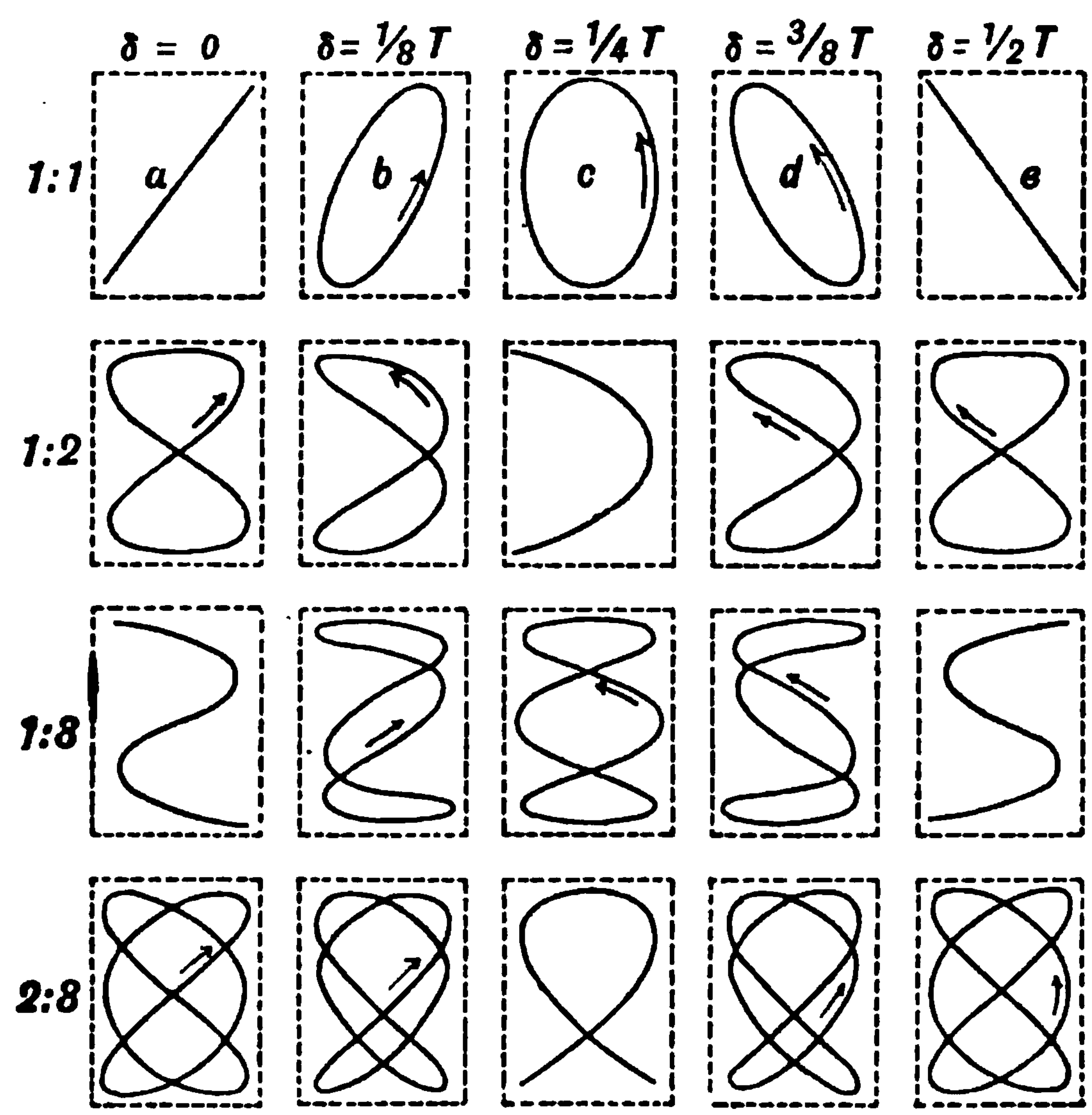


FIG. 145.

244. Composition of Two Simple Harmonic Motions at Right Angles with Periods in Simple Ratio.—Proceed as in the last case,

but divide the respective circles of reference into a number of equal parts proportional to the respective periods, so that the intervals in the two circles will be traversed in equal times. Fig. 144 illustrates the case where $T_1/T_2=1:2$, and the angular phase difference is 45° or the time phase difference $\delta=T_2/8$. Orbits corresponding to other phase differences and to the ratios $T_1/T_2=1:3$ and $2:3$ are shown in Fig. 145.

245. Lissajous' Figures.—Experimental illustrations of these curves were first obtained by Lissajous. His method was to reflect a beam of light from a mirror attached to one end of a tuning fork to a corresponding mirror on another fork vibrating in a plane at right angles to the first, and thence on a screen. The beam is displaced by both forks, and the spot of light on the screen describes the resultant path. Another method is to use a Y-pendulum, as shown in Fig. 146. If the bob vibrates in the plane of the paper, the effective length is PQ; if it vibrates at right angles to this plane, it is CQ. The periods in the two planes will, therefore, be different and independent. By properly adjusting the lengths PQ and CQ the bob may be made to describe the various Lissajous' figures. If the pendulum has a single support, $T_1=T_2$, and the bob will move in an ellipse, circle, or straight line, according to the difference of phase between two impulses given to it at right angles.

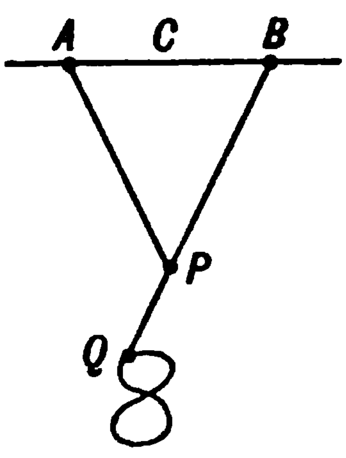


Fig. 146

A rectangular rod fastened at one end will vibrate transversely with a period depending on its thickness. If the diameters parallel to two sides are different, the respective periods of vibration will be inversely as the diameters. If drawn aside diagonally and released, the rod will not continue to vibrate in that direction, but the displacement will be resolved into two components parallel to the diameters. If the ratio of the periods is simple, the end of the rod will describe Lissajous' figures.

246. Waves due to Simple Harmonic Motion.—Consider a number of spheres of equal masses attached to each other by elastic connections, as in Fig. 147. If a transverse simple harmonic vibration is imparted to the first, the impulse will be transmitted to the others in succession. Suppose the phase difference between

the displacements of successive spheres to be one-eighth of a period. When *a* has completed one vibration, *b* has completed seven-eighths of a vibration, etc., while *i* is just beginning to move. The positions of the spheres will be at the projections on the vertical lines 1, 2, 3, etc., of the points 1, 2, 3, etc., of a circle of reference, with radius equal to the amplitude of the wave. If a smooth curve be drawn through these positions, it will give the wave form. It is evident that the abscissa of any point on this curve is proportional to the time required for the disturbance to reach that point, or to the phase angle, and its ordinate to the sine of the phase angle of the disturbance at the point. Such a locus is called a **harmonic curve or sine curve,** and gives the shape of a transverse wave when the medium executes simple harmonic vibrations. If the particles in Fig. 137

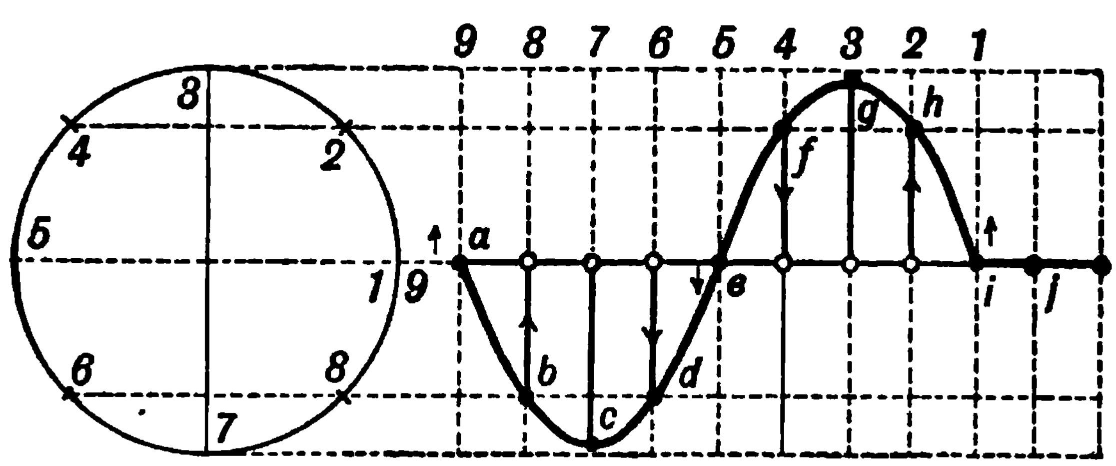


Fig. 147.

execute simple harmonic vibrations, the longitudinal wave will be of the same type, and may be represented by a sine curve.

The period and the **amplitude** of the wave are the same as those of the simple harmonic motion of any point in the medium. The **wave length** λ *is the distance between any two consecutive points in the same phase of displacement*, for example *a* and *i* (Fig. 147). If v is the velocity of propagation of the wave, $vT=\lambda$, since λ is the distance traversed by the wave during a complete vibration of the "source," sphere *a*. If $n=1/T$ is the frequency of vibration, $v=n\lambda$, the length of the train of waves sent out in one second.

The displacement in a longitudinal wave presents the same aspect if looked at from any direction in a plane at right angles to the direction of propagation. This is not the case with the transverse waves represented in Figs. 136, 147, for the vibrations

will be in the line of sight if viewed in the plane of vibration, and at right angles to the line of sight if viewed normally to this plane. These transverse waves have a sort of polarity, therefore, and are said to be *plane polarized.*

Transverse waves may be set up in a cord or longitudinal waves in a spiral spring by fixing one end and attaching the other to a vibrating tuning fork. The amplitude of the waves in such cases may be much greater than that of the fork.

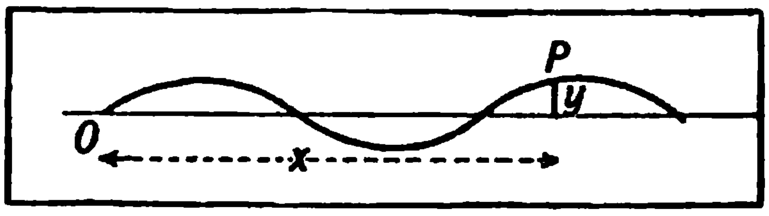


FIG. 148.

If a beam of light be reflected from a mirror attached to the end of a vibrating fork, and again reflected to a screen from a revolving mirror, the harmonic curve will be traced on the screen by the spot of light. Persistence of vision will cause the path to appear continuous.

A permanent record of such curves may be made by causing a bristle attached to the end of a tuning fork to trace its path on the smoked surface of a piece of glass which is moved past the fork at a uniform rate v.

The coördinates at the time t of a point P on the sine curve, with respect to the origin O, are evidently (Fig. 148).

$$x = vt$$
$$y = r \sin \frac{2\pi}{T} t$$

Eliminating t,

$$y = r \sin \frac{2\pi}{T} \frac{x}{v} = r \sin \frac{2\pi}{\lambda} x$$

This is the equation of a sine curve repeating itself at intervals of $x = \lambda$.

If $y = r \sin 2\pi t/T$ is the harmonic displacement at a given point, the disturbance will reach a point at a distance x in the time $t_1 = x/v$; the disturbance at the point x at the time t will have the phase

$$\frac{2\pi}{T}\left(t - t_1\right) = \frac{2\pi}{T}\left(t - \frac{x}{v}\right)$$

and

$$y = r \sin \frac{2\pi}{T}\left(t - \frac{x}{v}\right) = r \sin 2\pi \left(\frac{t}{T} - \frac{x}{\lambda}\right)$$

This is the equation of wave motion. At a given time t, say $t = 0$, it gives the instantaneous picture of the wave train as a sine curve. At a given point, say $x = 0$, it represents the simple harmonic vibration of the medium at that point.

247. Superposition and Interference of Waves.—If two or more trains of waves are superimposed, each will give rise to independent displacements of the medium. The resultant effect may be obtained, therefore, by plotting each train of waves on the same axis, with relative displacements corresponding to their phase differences, and adding the ordinates. It is convenient to express the phase differences in terms of wave-length. If, for example, one wave starts half a period later than another, it should be plotted with its front half a wave-length behind that of the first. In Fig. 149, *A*, *B*, *C*, the full line represents the resultant of two waves of the same length and with phase differences of 0, $\lambda/4$, and $\lambda/2$, respectively. In the last case the resultant effect

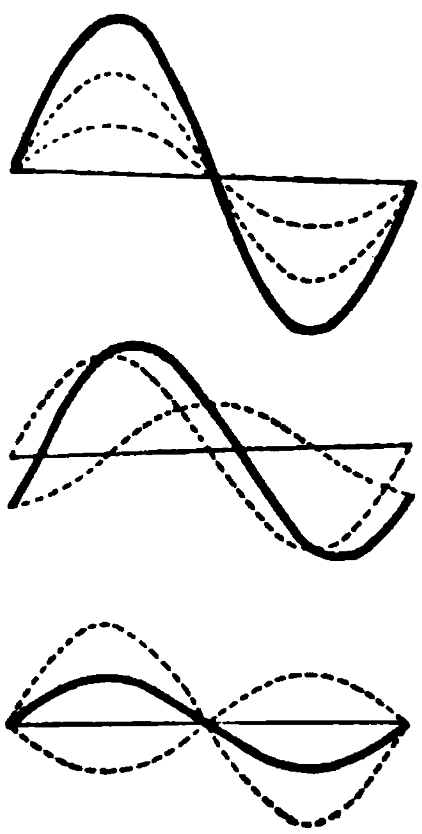


FIG. 149.

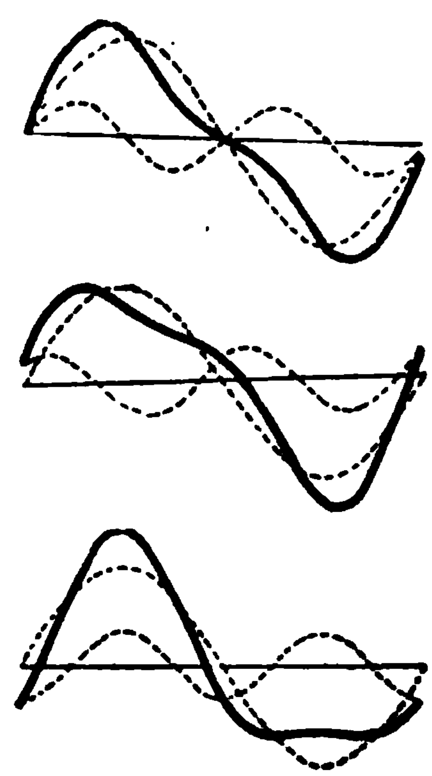


FIG. 150.

is zero if the amplitudes are equal. The modification of amplitude due to the superposition of waves is called **interference**. It is evident that the length of the resultant wave is the same as that of its components, and that it is a harmonic curve if they are harmonic curves.

248. Complex Waves.—Waves of different lengths may be combined in the same manner. If the lengths are in simple ratio to one another, all the resultant waves in a train will be of the same form, but this form will vary with the phase difference, and will not be a sine curve. This is illustrated by Fig. 150, *A*, *B*, *C*, which shows the resultant of two waves of lengths in the ratio 1 : 2, and having different phase relations. Fig. 151 illustrates

the case where the lengths are as 1 :3 and the phase difference zero.

If the components have lengths which are not in simple ratio, successive waves will not be of the same shape, as the length of the longest wave will not be a common multiple of the lengths of the component waves. If there are only two components, however, with frequencies n_1 and n_2, one wave will gain its own length on the other in $1/(n_1-n_2)$ second, and the wave train will consist of similar groups repeating themselves n_1-n_2 times a second. The length of each group will be the least common multiple of the lengths of the components. Fig. 152 shows the effect of superimposing two trains of waves of lengths having the ratio 3 :4. The graphical representation of "beat" waves in sound would resemble this figure (§602). Such forms may be obtained experimentally by the optical method for obtaining sine curves described in §246, the beam of light being reflected successively from two forks vibrating in the same plane, and giving beats, and from a rotating mirror to a screen.

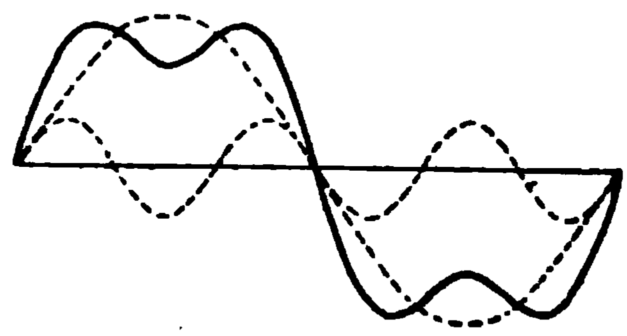


FIG. 151.

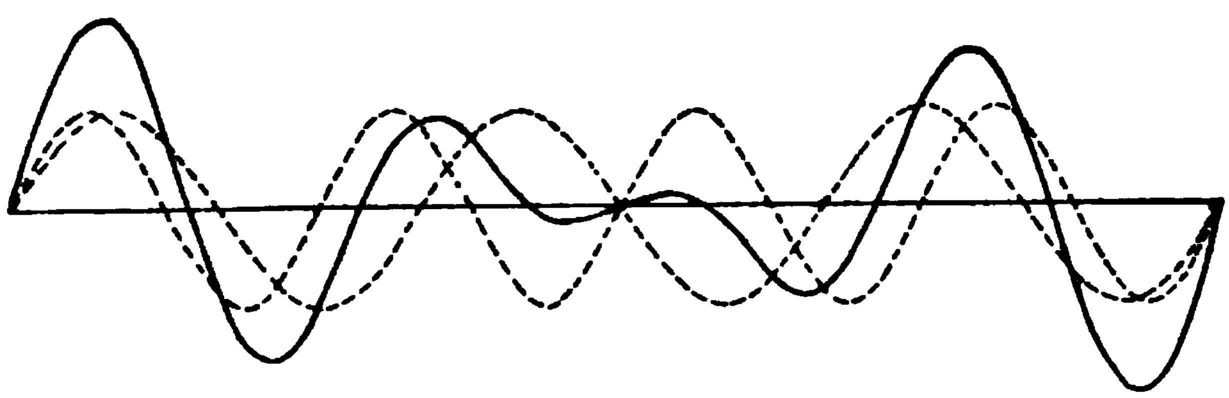


FIG. 152.

The displacements of the medium may be the resultant of two displacements at right angles. The example of water waves has already been mentioned. If one end of a cord be attached to the end of a rectangular rod vibrating transversely parallel to each of its diameters, the end of the rod will describe Lissajous' figures (§245) and each element of the cord will do the same. If the two diameters of the rod are equal, each element of the cord will move in a circle or ellipse in a plane transverse to its lengh,

but the phases will differ from point to point, so that at a given instant the cord will have the shape of a corkscrew. Such a wave is said to be *circularly* or *elliptically polarized.*

249. Fourier's Theorem.—The illustrations given show that various complicated forms may be obtained by the addition of simple harmonic waves of different lengths and phases, and that these waves will be of persistent form if the periods of the components are simple fractions of the periods of the longest component. Fourier proved that any periodic disturbance or wave form of permanent type could be represented as the summation of a number of simple harmonic terms of the form

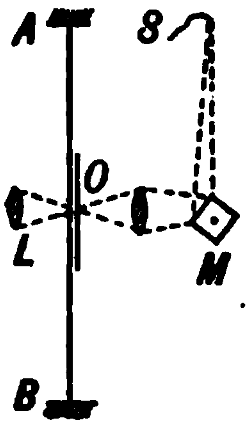


FIG. 153.

$$x = r_1 \sin \omega t + r_2 \sin 2\omega t + r_3 \sin 3\omega t + \cdots, \text{ etc.,}$$

the periods and wave-lengths of the components having the ratios 1, $\frac{1}{2}$, $\frac{1}{3}$, $\frac{1}{4}$, etc. Fig. 151 shows that the resultant is approaching a rectangular form, which may be finally attained by adding shorter waves.

The forms of complex waves may be projected by the following device (Fig. 153): A screen with a slit opening at O is placed in front of a horizontal stretched wire AB, which is illuminated by the lens L. An image of the segment opposite the opening may be thrown on a screen S after reflection from a rotating mirror M. If the wire is at rest the image of the illuminated segment will be drawn out in a dark straight line on the screen. If the wire vibrates in a vertical plane the images of the segment in its successive phases as the wave passes O will be laid off end to end on the screen, giving the actual form of the wave passing the opening.

250. Velocity of a Wave on a Cord.—Let the wave be supposed to be moving toward the left with a velocity v. It will simplify the problem without essentially changing it, if we now suppose the cord to be given a velocity v toward the right. The wave will then stand still and every part of the cord, as it comes to the wave, will pass through it with velocity v. This, in fact, is what may often be noticed in the use of a chain hoist. If the chain be started in rapid motion (there being no load on the lower pulley), a bend impressed on the chain will sometimes remain stationary for a short time, and, if the chain be suddenly arrested, the bend will move off in the opposite direction with (approximately) the speed which the chain had. The *relative* velocity depends only on the mass and tension of the chain.

Now let QR be a small part of the wave, its length l being so short that it may be regarded as an arc of a circle (the circle of curvature). Draw tangents TQ and TR and complete the par-

allelogram $QTRS$. The velocities at Q and R may be represented by QT and TR. As each part of the cord passes from Q to R in time t it will have an acceleration, a, toward the center of curvature, that is, in the direction of the diagonal TS. Since TS represents a velocity which, added to QT, gives QS or TR, it represents the change of velocity, at, in the time t. Hence

$$\frac{TS}{QT}=\frac{at}{v}$$

The only forces that act on the part QR of the cord are the equal forces, F, at its ends due to the tension in the cord. These may be represented by TQ and TR and their resultant, represented by TS, is the force that causes the central acceleration

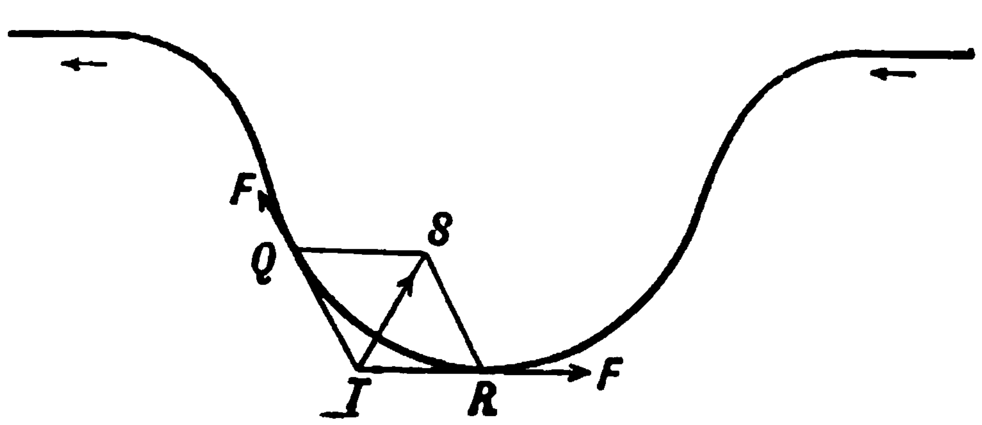


FIG. 154.

of the part QR of the cord. If the mass of unit length of the cord is m, the mass of the part QR is ml. Hence

$$\frac{TS}{QT}=\frac{mla}{F}$$

Equating the above values of TS/QT and noting that $l=vt$, we get

$$v=\sqrt{\frac{F}{m}}$$

(It is evident that belting traveling with this velocity will exert no pressure on a pulley. See §47.)

251. Velocity of Elastic Waves.—It might be expected that the velocity of waves in an elastic medium would depend upon the elasticity, which determines the rate at which an impulse is transmitted from one element to another (in a perfectly rigid and incompressible medium the effect would be instantaneous), and the density, which exercises a retarding influence, on account of the inertia of the displaced elements. The derivation of the exact relation between the velocity, the density ρ, and the coefficient of elasticity E is in some cases mathematically difficult, but the general

form, at least, is readily obtained from a consideration of the dimensions of the quantities involved (§154). If the velocity depends solely on ρ and E, we may write $v = kE^x\rho^y$, where x and y are unknown powers, and k a factor of proportionality. Substituting dimensional expressions for the quantities (remembering that E is force per unit area and ρ is mass per unit volume), we have

$$(v) = \frac{(L)}{(T)} = \left(\frac{ML}{T^2L^2}\right)^x \left(\frac{M}{L^3}\right)^y$$

By inspection we find with respect to T that x must be $\frac{1}{2}$. To make M disappear from the right-hand side, y must be $-\frac{1}{2}$. Therefore

$$v = k\sqrt{\frac{E}{\rho}}$$

The exact relation may be easily found in some simple cases. Suppose the front of the disturbance in a longitudinal wave in a medium of unit cross-section to be at A (Fig. 155) at one instant and at B a short time t later. The velocity of the wave is, therefore, $v = l/t$, where $l = AB$. An imaginary plane A in the medium is displaced to D, a distance x, by compression. If l is very small, the density of the substance is practically uniform between D and B, and the center of mass of the element is displaced from C to C', a distance $x/2$.

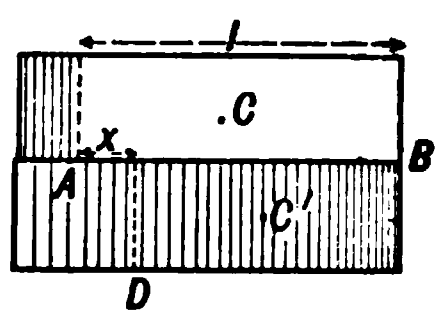


FIG. 155.

The average velocity of the center of mass is $x/2t$ and its final velocity x/t. The final force acting on the element is Ex/l, where E is Young's modulus in the case of a solid, or the modulus of elasticity of volume in case of a gas. The average force is half of the above. Equating the work done by this force to the acquired kinetic energy due to the motion of the center of mass, we have

$$\frac{Ex}{2l}x = \frac{1}{2}\rho l\frac{x^2}{t^2}; \text{ therefore, } \frac{l^2}{t^2} = v^2 = \frac{E}{\rho}$$

This applies to longitudinal waves in a wire or rod or to sound waves to any medium.

252. Reflection of Waves.—When a transverse wave reaches the fixed end of a cord, the displacement is immediately reversed in direction by the elastic reaction of the fixed end. The wave is, therefore, reflected with reversal of phase of displacement, as shown in

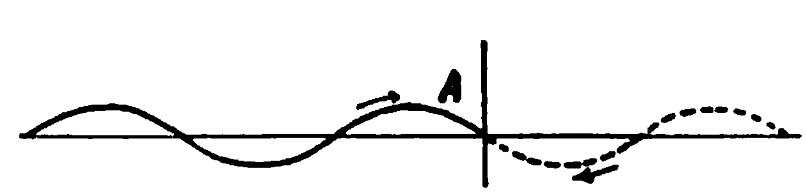


FIG. 156.

Fig. 156. Apparently the incident wave has disappeared through the end, while a wave of opposite displacement has entered, traveling in the opposite direction, and at every

instant exactly neutralizing the displacement of the end which would be caused by the incident wave if the end were free. When a continuous train is reflected, the effect is as though a train of indefinite length had been cut in two when a wave-front reaches A, the fixed point (Fig. 156), and the waves to the right immediately reversed in direction, while the incident waves continue their motion unchanged; or as though a train of incident waves were traveling through a mirror, while their inverted images proceed out of it in the opposite direction.

If one end of the cord is free, when the wave reaches that point, the end, having nothing beyond to restrain it, has an outward displacement twice as great as though the cord were continuous, and it will, therefore, immediately start a wave of the same phase in the reverse direction. After half a period of vibration it will return through the resting point in the opposite direction, and will start a backward wave with phase opposite to that of the incident wave. It is as though a train has been cut in two when a crest is at the free end B (Fig. 157), and the right-hand section immediately reversed; or as if an advancing train were passing through a mirror while its erect image emerged from it.

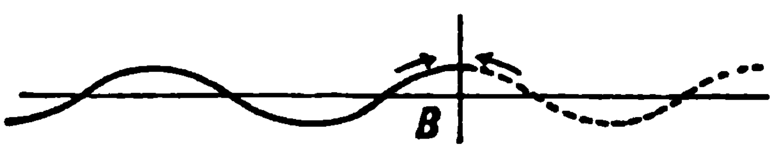


FIG. 157.

To show in another way the difference between reflection at a free end and at a fixed end, suppose that the part of the direct wave-train that first reaches B and begins to be reflected is as represented in Fig. 158. If we compare this with Fig. 156, it is seen that in reflection at a free end, as compared with that at a fixed end, *there is a delay of half a period* in the reflection of the wave of opposite phase, as though the right-hand section in Fig. 156 were held at rest for half a period before starting in the negative direction.

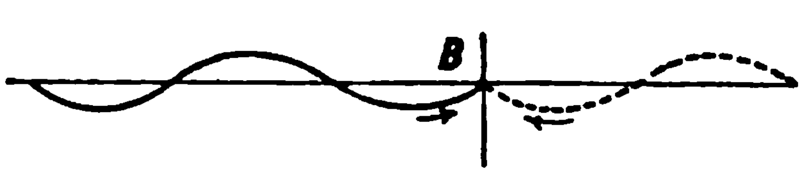


FIG. 158.

Reflection of longitudinal waves may be illustrated by the conduct of a row of elastic pendulums of the same size, as shown in Fig. 159a, the last resting against a fixed obstacle. If a is drawn aside and released, it will impart an impulse to b, this in turn to c, etc., and a compression wave will travel to the other end

of the row; *g* cannot move, but will be compressed, and through its elastic reaction it will almost immediately start a compression wave in the opposite direction. When this wave reaches the free end, *a* will fly out without restraint, leaving a rarefaction behind it; or, if elastically connected with *b*, it will at once send back a rarefaction wave. In any event, after executing half a vibration it will swing back through its equilibrium position and reflect a compression wave to the right.

If there are two rows of elastic pendulums of different masses, as in Fig. 159*b*, displacements will be immediately transmitted across *A*, no matter in which direction the wave is moving, but the wave will be also partially reflected. If the wave travels to the right, reflection of *a* at *A* will be immediate; if it travels to the left, the more massive sphere *b* will continue after impact to move to the left, and will return through its resting point, to send a wave back to the right, at the expiration of half its period of vibration.

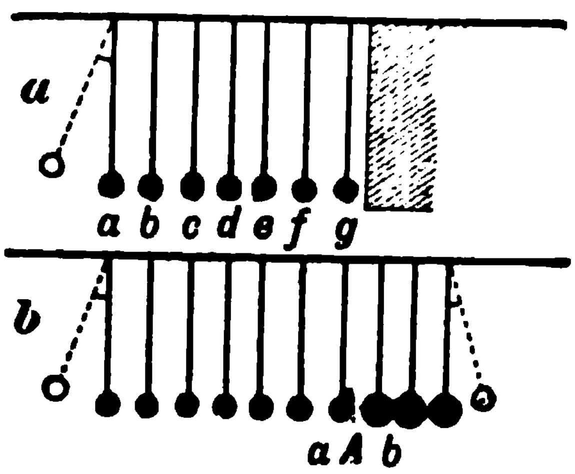


FIG. 159.

The cases mentioned illustrate the general principle that *the displacements in a medium have a minimum amplitude at a fixed or constrained boundary; a maximum amplitude at a free boundary or one with diminished constraint.* Important illustrations of this principle occur in cases where waves pass from a light to a dense medium or *vice versa* (§§609, 686).

253. Stationary Waves.—Consider a train of waves in a cord moving to the right, while a similar train (reflected or independent) moves to the left. Interference will take place, and the resultant displacement of the medium at a given point and time will be the sum of the individual displacements. Plot the positions of the waves at successive instants (say at intervals of an eighth of a period). If the incident train is represented by a light line, the reflected train by a dotted line, and the resultant by a heavy line (Fig. 160), it will be seen that there are always points of zero displacements *N* (or of minimum displacement if the amplitudes are unequal) at intervals of half a wave-length, where the waves always meet in opposite phases.

Half way between these points, at L, the waves will always meet in the same phase, and the displacement will be a maximum. The former positions are called **nodes**, the latter **loops** or **antinodes**. Between the nodes the medium oscillates back and forth, the direction of the displacements being opposite in adjacent segments, so that at any instant the cord has a more or less sinuous shape, except at intervals of half a period, when it passes through the undisturbed straight position (Fig. 137*d*). The same conclusions apply to longitudinal waves. Disturbances of this sort are called **stationary** waves. It is evident that when these arise from the interference of incident and reflected waves there must be a node at a fixed or constrained boundary, a loop at a free or unconstrained boundary.

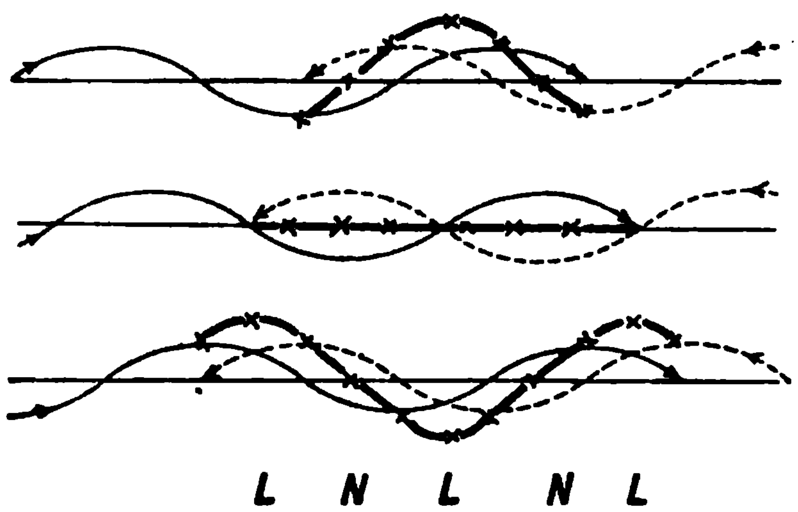


FIG. 160.

Fig. 161 is the graphical representation of stationary waves of longitudinal type. The displacements have just begun to return from maximum elongation, from the full to the dotted line. This indicates that the particles to the left of N_1 and those to the right of N_2 are moving in the negative direction, while those between N_1 and N_2 are moving in the positive direction. Consequently the particles on opposite sides of N_2 are approaching that point, while those on opposite sides of N_1 are receding from it. At N_2 there will be a condensation, at N_1 a rarefaction. After half a period conditions will be reversed. In the neighborhood of L_1, L_2, however, the particles are moving in the same direction with approximately the same velocity, so that their relative positions are only slightly changed. It follows that *at the nodes there are the greatest variations of pressure, and the least motion; at the loops, the smallest variations of pressure and the greatest motion.*

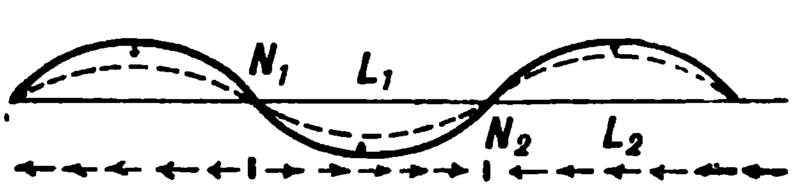


FIG. 161.

254. Waves in a Liquid.—Some of the most interesting properties of wave motion may be illustrated by waves on the surface of

a liquid, such as water. The initial displacement may arise from differences of level caused by some external force, such as the impact of a pebble, winds, etc. The effect of gravity, of fluid pressure, and of surface tension is to restore the original level, but, on account of their inertia, the particles are displaced beyond their equilibrium positions, just as in the case of vibrations of a liquid in a U-tube. Horizontal as well as vertical displacements must occur, as in the case of the liquid in the bend of the U-tube. There is, therefore, a longitudinal as well as a transverse component. These displacements are simple harmonic, because the resultant pressure on an element is proportional to its vertical displacement from the undisturbed surface. We have seen (§236) that on a crest the element moves forward, in the hollow backward, in intermediate positions both vertically and horizontally. Fig. 162 shows the positions and directions of rotation

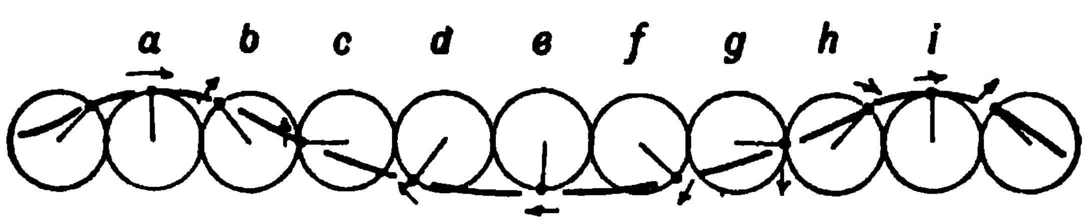


FIG. 162.

of a number of particles originally at rest on the surface in the positions under *a*, *b*, *c*, etc., the phase difference between successive displacements being an eighth of a period. Particle *a* is subject solely to a downward acceleration, particle *e* to an upward acceleration; particles *c* and *g* are subject solely to horizontal accelerations, due to the lateral pressure, as they are in the horizontal plane of equilibrium. We thus find that there is a difference of phase of a quarter period between the vertical and horizontal accelerations, in accordance with the observed fact that the disturbed elements move in elliptic or circular orbits. It is evident that the wave form is not a sine curve.

The expression for the velocity of liquid waves is complicated and cannot be derived here. It is sufficient to say that large waves are maintained by gravity alone, and that the velocity is independent of the density of the liquid, as the force acting is proportional to the weight of the displaced elements, and hence will produce the same acceleration, whatever the density. The velocity increases with the wave-length, so that one may frequently see a train of long water waves sweeping through a train

of shorter waves and leaving them behind. When the liquid is shallow, the velocity diminishes with the depth. The very small waves are maintained by surface tension alone, so that they are analogous to transverse waves in an elastic membrane. In the case of these waves the velocity increases as the wave-length diminishes, and is also dependent upon the density and the surface tension. Such waves are called **ripples**.

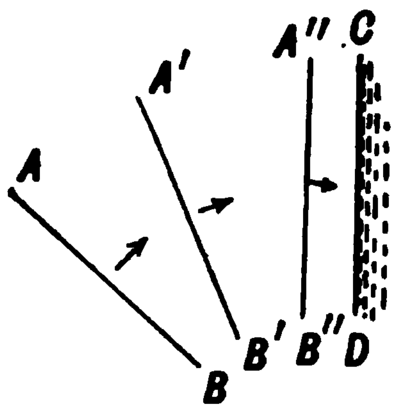


FIG. 163.

255. Refraction of Waves.—Waves move more slowly in shallow than in deep water. Hence if the front AB of an ocean wave moving in the direction of the arrow (Fig. 163) approaches a beach CD, the nearer end, B, of the wave will be retarded more than A, being in shallower water. The wave front will swing around into the successive positions $A'B'$ and $A''B''$, and will finally become parallel to the shore line. This change in direction due to change in velocity is called **refraction**. Similar effects are, we shall see, shown by other waves, such as sound and light, when they pass from one medium to another in which they travel with different velocity.

256. Propagation and Reflection of Ripples.—Experiments with ripple waves may be shown by the following arrangement. A shallow wooden box with a glass bottom, about two feet square, is mounted on legs like a table, carefully leveled, and partly filled with water. Light may be projected upward through the bottom from an arc light placed beneath the box, or by reflecting a divergent beam of sunlight upward by an inclined mirror. Ripples on the surface will by their lens effect change the distribution of light on the ceiling so that the motion of each ripple may be followed.

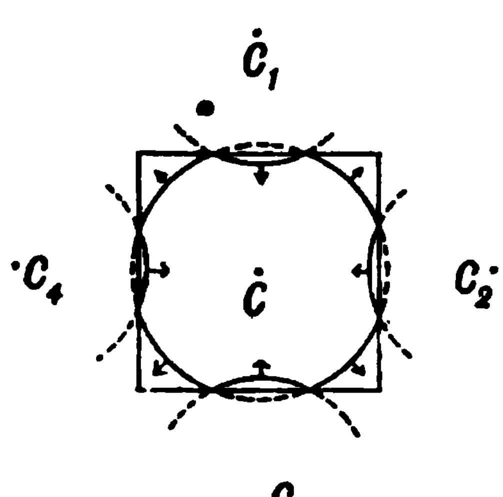


FIG. 164.

If the middle of the surface is touched with a nail, a circular ripple will diverge from that point. If the surface were larger, this wave would at a later time occupy the position of the circle (Fig. 164), but it will be in part reflected from the four sides. The reflected segments are exactly like the missing segments of the

outgoing wave, reversed in direction. These reflected waves have centers at C_1, C_2, C_3 and C_4, the "images" of the source C, which are evidently at the same distance from the walls as the source itself, since C and the other centers of curvature are symmetrically situated with respect to the walls. These reflected waves will cross each other and be subject to repeated reflections ("multiple reflection"), their curvature all the while decreasing, until we have a rectangular system of straight ripples.

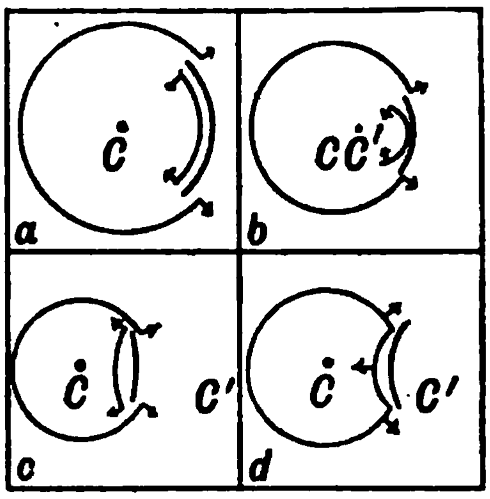


FIG. 165.

If a circular wave strikes a bent sheet of metal of the same curvature as the wave, the latter will be reflected without change of curvature, converge to its starting point, and diverge from it on the opposite side (Fig. 165*a*). If the strip has a greater curvature than the wave, the edges of the latter will be first reflected, so that its curvature is increased (*b*). It will converge to C', a "real image" or "focus." If the strip is concave with less curvature than the wave (*c*) the latter may be made divergent, with a "virtual" image at C'. If the strip is convex toward the wave, the reflected wave will always diverge from a virtual center behind the mirror (*d*.)

If the surface is touched with a long straight strip of metal a straight ripple will be produced. If this strikes a screen with a small slit in it (Fig. 166*e*) the disturbance will pass through this hole and set up a semicircular wave on the other side. The remainder of the wave will be reflected as a straight line.

If a number of nails are driven at equal distances through a strip of wood and dipped into the water, circular waves will diverge from the points of contact. At a little distance these wavelets will blend into a straight ripple corresponding to their common tangent (*f*). At other points the ripples cross each other in all phases, and their effect will vanish because of interference. We may, therefore, consider that a linear wave front is due either to a continuous linear disturbance or to a number of neighboring point dis-

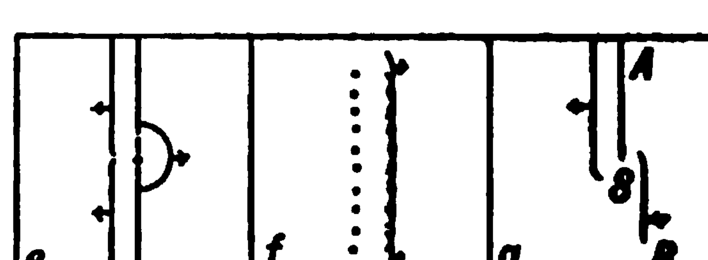


FIG. 166.

turbances, each sending out circular waves. In (*e*) for example only the point in the opening is effective for transmission. The latter conception is often useful (§639).

If a screen *S* projects part way across the tank (*g*), the portion *AS* of an incident wave will be reflected; the remainder *SB* of the wave will pass the screen. It will be noted that the end of the transmitted wave front will bend into the shadow of the screen, and the end of the reflected wave will bend into the region formerly occupied exclusively by the other half of the wave. *S* is apparently a center of disturbance for both these waves. This effect is called **diffraction.** By noting the resemblance of the ends of the waves in this case to those in the preceding case (*f*) the explanation will be made clear.

257. Refraction of Ripples.—Advantage may be taken of the fact that the velocity of water waves diminishes with the depth to illustrate refraction. On the bottom of the tank lay a piece of thick glass, so that the water over it is about one-fourth as deep as elsewhere. A linear ripple is started by touching the surface with a strip of metal. On reaching the edge of the glass plate

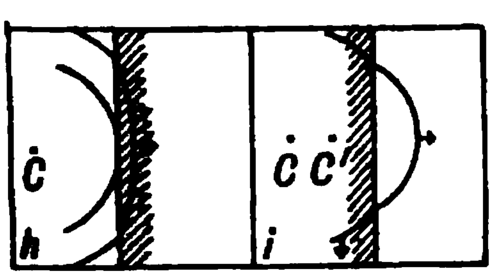


FIG. 167.

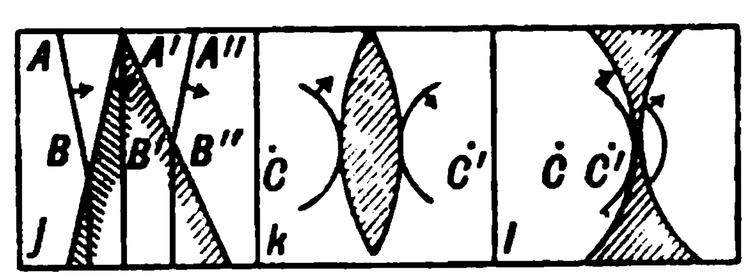


FIG. 168.

the end *B* is retarded and the wave will swing into the position *A′B′*, as in Fig. 163.

If the incident wave is circular, the middle will be more retarded than the edges if the wave comes from the deeper water, and the curvature of the wave will be diminished (Fig. 167*h*). If the wave travels from the shallow region, the contrary will be the case (*i*). The centers of curvature or "images" of the source will be at *C′* (outside the tank in *h*).

If a prismatic sheet of glass is laid on the bottom (*j*) a linear wave front *AB* will be rotated both in approaching and leaving, and the final direction will be *A″B″*. If pieces of glass with convex or with concave edges, like sections of lenses, are laid on the bottom, the center of a passing circular wave will be more

retarded than the edges in the first case (k), less retarded in the second (l), resulting in changes of curvature. The "images" of the source will be at C'.

258. Interference of Ripples.—If two nails simultaneously touch the water at different points two circular waves will be set up, which will cross and interfere with each other. They pass so quickly, however, that it is difficult to observe them. Better results will be secured if a continuous series of waves can be produced, and still better results if there is a system of stationary waves. A very satisfactory method of securing this result is to put mercury in a circular glass dish at least four inches in diameter, and maintain periodic disturbances at the center by a glass fiber attached to the vibrating prong of a tuning fork Continuous trains of circular ripples will diverge from the center, while reflected circular ripples will converge

Fig. 169.

toward that point The result will be a system of circular stationary waves, as illustrated in Fig 169. They may be projected on a screen by reflected light, and made more distinct by using a lens.

If two glass fibers are attached to the fork near each other, two trains of waves will be maintained, and each will form its own system of stationary waves. At all points on the surface where the outgoing waves meet each other in the same phase (that is, where the difference of the respective distances to the two sources is zero or any whole number of wave lengths), the waves will reinforce each other. In regions where they meet in opposite phases (the differences of path being some odd multiple of a half wave length), they will destructively interfere with each other. Along certain lines, therefore, there will be no disturbance by either outgoing or reflected ripples (Fig. 482). Between these lines segments of the stationary waves appear, as shown in Fig. 169.

259. Energy and Intensity of Waves.—The energy of a vibrating body is proportional to the square of its amplitude (§61). Each vibrating element of mass in a medium traversed by waves will, therefore, possess energy proportional to the square of its amplitude, and this energy will flow forward with the advancing waves. The **intensity** of waves in a given region is defined as being proportional to the amount of energy passing per second through unit area at right angles to the direction of propagation; hence the intensity is proportional jointly to the square of the amplitude and the velocity of the waves. In a viscous medium, such as molasses or lead, they rapidly decay in amplitude and disappear, owing to the absorption of energy by internal friction. This effect is known as **damping**. Fig. 138 represents the form of a damped train of waves. If there is no such loss the same quantity of energy will persist in a given wave, no matter how far it travels, or how the dimensions and form of the wave front may change. If such waves travel in a wire or any other channel of constant cross-section the intensity will be independent of the distance from the source, as the wave front will remain of constant area. This is illustrated by the transmission of sound waves through a speaking tube or of light waves in a parallel beam. In the case of circular waves on a surface, a constant amount of energy will remain in a wave of circumference which increases directly as the distance from the source; hence the intensity must vary inversely as the distance, and the amplitude inversely as the square root of the distance. In the case of spherical waves, the energy will remain constant within a spherical shell of the thickness of one wave length and with surface increasing as the square of the distance. If E is the energy emitted from the source per second, and if r_1 and r_2 are the radii of the wave at different distances, and I_1 and I_2 the corresponding intensities.

$$E=4\pi r_1^2 I_1=4\pi r_2^2 I_2 \quad \therefore \frac{I_1}{I_2}=\frac{r_2^2}{r_1^2}$$

Hence the intensity varies inversely as the square of the distance from the source, and the amplitude inversely as the distance.

References

FLEMING'S *Waves and Ripples in Water, Air, and Ether* is an excellent popular description of all kinds of waves.

EDSER'S *Light*, chapters on Wave Motion.

DANIELL'S *Principles of Physics*, chapters on Wave Motion.

WOOD'S *Physical Optics*, Ch. 3 and 4, gives an interesting account of the photography of sound waves.

Problems

Simple Harmonic Motion.

1. A mass of 196 grams is suspended by a rubber band of such elasticity that an additional weight of 5 grams will stretch it 1 cm. It is extended 1 cm. and released. Find the period, and the displacement, velocity, and acceleration 9 seconds after it passes upward through its resting point.
   Ans. $T=1.256$ sec.
   $x=\sin 59°.6=0.862$ upward.
   $v=2.45$ cm./sec. upward.
   $a=21.55$ cm./sec.^{2} downward.
2. Water or mercury in a U-tube is disturbed. Show that the liquid executes a simple harmonic motion of period $T=2\pi\sqrt{l/2g}$, where l is the length of liquid from surface to surface around the bend.
3. Compound two simple harmonic motions of same period and in same plane with amplitudes 3 and 2 and with phase difference of one-sixth of a period. *Ans.* $R=4.36$.
4. Compound two simple harmonic motions at right angles with periods in the ratio 3 : 5 and with phase difference zero.

Waves.

5. Compound three trains of waves of lengths in the ratios 1, $\frac{1}{2}$, and $\frac{1}{3}$ and of amplitudes 3, 2, and 1, starting in the same phase.
6. Compound two trains of waves of lengths 5 and 4 and of equal amplitudes.
7. A copper wire ($\rho=8.8$) 1 square mm. in cross-section is subject to a tension of 88,000 dynes. With what velocity will a transverse wave travel in it? *Ans.* 1000 cm./sec.
8. With what velocity will a longitudinal wave travel in the same wire? *Ans.* 350,000 cm./sec.

HEAT

By Charles E. Mendenhall, Ph. D.

Professor of Physics in the University of Wisconsin

INTRODUCTION

260. Early Ideas.—The preceding sections have dealt with physical changes involving, in general, motion and changes in motion of bodies as a whole. We have now, however, to consider changes in physical condition which do not involve obvious changes in motion, of which the most common are changes in hotness or coldness and changes in state, that is, melting or boiling. The sense of touch is the first and simplest means of distinguishing hot from cold bodies, and by it we can roughly arrange bodies in the order of their hotness, deciding that A is hotter than B, B than C, etc. But the sense of touch is found to be neither reliable nor delicate enough to be used as a *measure* of degrees of hotness, and, moreover, a limit of hotness or coldness is very soon reached beyond which the touch sense cannot be directly applied. A purely physical basis of measurement (§264), depending on the properties of bodies, is therefore adopted which agrees with the sense of hotness as far as they can be compared. When measured in this definite physical way the hotness of a body is called its *temperature*, the scale of measurement being so chosen that hotter bodies have higher temperatures.

It is found that increase in temperature of a given body can be produced by various common causes, such as contact with or exposure to fire, contact with a hotter body, and friction, as, for example, rubbing ones hands together. The causes which will produce increase in temperature will also, under proper conditions, produce melting or boiling and various other physical changes, of which increase in size is the most common and obvious. On account of these common causes it was most natural to group together the various effects referred to as they became known and attribute them all to the passage, into or out of bodies,

of a substance called *caloric* or *heat*, the presence or absence of which accounted for all of these related phenomena. According to this theory, heat was a material substance, but one which could not be weighed or detected by any ordinary physical method. On the basis of this hypothesis fairly consistent explanations were given for many common facts. For example, the temperature of a body was said to depend on the amount of caloric it contained and upon its natural capacity for caloric, which in turn depended upon its physical state, as, for instance, the state of subdivision. A given amount of matter in powdered form was thus supposed to have a less capacity for caloric than the same quantity in larger pieces. Thus the rise in temperature produced by rubbing two bodies together was explained as being due to the abrasion of the material, its capacity for caloric being thereby reduced and a certain proportion of its caloric set "free," and its temperature correspondingly raised. According to this idea, the entire amount of caloric set free should, under given circumstances, be proportional to the entire amount of material abraded.

261. Heat and Work.—The first serious question of the truth of the caloric theory was raised in 1798 by Count Rumford who, in experiments carried out in Munich upon the caloric developed in the boring of cannon, used a blunt borer which cut very little material, and arranged matters so that the heat generated raised the temperature of a considerable quantity of water which was made to boil "without fire." From these experiments he concluded that the amount of caloric developed was not at all proportional to the amount of abrasion but was, at least approximately, proportional to the amount of *mechanical work* required to do the abrading. In the following year Sir Humphrey Davy performed the similar but more striking experiment of melting ice by rubbing two blocks of it together, the temperature of the ice as a whole being below freezing, and again it was concluded that the melting was due to the transmission of *motion* to the ice molecules. From this time on the idea that heat could be produced from mechanical motion and *vice versa*, or, as it is put to-day, *that heat is a form of energy*, was gradually accepted. But it was nearly 50 years before the full significance of this new point of view was appreciated and careful measurements were made by

Joule and others of the amount of work equal to a given amount of heat. This idea that heat is a form of energy, together with the ideas of the kinetic theory of gases (§227), and the conception of the molecular structure of matter suggested by chemical and radioactive (§580) investigations, unite to give the present *molecular* or *kinetic theory of heat.*

262. Molecular Theory.—According to this point of view matter consists of units or parts called molecules, which are composed of smaller units of the *elements* (oxygen, hydrogen, iron, etc.), called atoms, these in turn containing still smaller units, namely elementary charges of negative electricity called electrons (§159) and probably a nucleus or center of positive electricity. Very little is known as to the structure of atoms, but the electrons in the atoms undoubtedly move about or vibrate very considerably, possibly somewhat as planets move about the sun, while the atoms move about inside the molecule, and molecules move inside the mass of matter, with great freedom when the matter is gaseous, with less freedom when it is liquid or solid (§§157-161). It is also possible under various conditions to have electrons existing more or less independent of atoms as "free" electrons or negative electric charges, the atoms which have lost electrons then having a positive electric charge, and being ready to capture any other electron which happens to come near enough; free electrons are characteristic especially of metals. Broadly speaking, the addition of heat energy to a body either increases the (kinetic) energy of motion of its molecules or increases their (potential) energy of position, as when melting or boiling occurs.

Considering this more in detail we see that all of the possible motions of molecules, atoms, and electrons would involve kinetic energy. Moreover, it is evident that changes of position of molecules, atoms, and electrons with respect to each other, against whatever forces, electrical or "chemical," may exist between them, would involve doing work against these forces, that is, changes in potential energy. Hence we can see that, when heat energy is added to a body, it may appear:

1. As an increase in the kinetic energy of motion of the molecules and free electrons.

2. As an increase in the potential energy of the molecules with respect to each other, in case the average distance separating them is increased.

3. As an increase in kinetic and potential energy of atoms and electrons inside the molecules.

This analysis of the possible changes in what is called the *internal energy* of bodies should be kept in mind throughout the study of heat, which will be found to be largely a study of the effects of changes in internal energy upon the condition and properties of matter.

THERMOMETRY

263. Standard Scale of Temperature.—We shall throughout use the term *temperature* to mean a quantity which we are now to define and which can be measured for any body at any time. Differences of temperature are to agree with our ordinary ideas of differences of hotness or coldness, so far as the two can be compared. The scale of temperature which we shall adopt is the international legal standard and is based upon the effect of increase in hotness upon the pressure of hydrogen. *Changes of temperature are defined as being proportioned to the corresponding changes of pressure in a constant mass of hydrogen confined at constant volume. This is called the hydrogen constant volume scale.* To measure the temperature of a body, for example, of a barrel of water, the vessel containing the hydrogen would be held in the water and the pressure of the hydrogen measured. But before temperature can be expressed as a *number*, we must have a unit in which to express it and we must also agree on a reference point or "zero" from which it is to be measured. The ordinary zero (reference point) is taken as the temperature of a mixture of pure ice and water when the pressure on the water surface is 1 atmosphere, while the degree is fixed by adopting a second standard point, the temperature of boiling water when the pressure is 1 atmosphere, which is defined as $+100°$ or $100°$ above zero. The *degree* is then such a change in temperature as will

Fig. 170.—Constant volume gas thermometer.

produce $\frac{1}{100}$ the change in pressure which is observed when the hydrogen is heated from the freezing- to the boiling-point of water. These specifications define the *Centigrade Zero* and Centigrade degree, which are universally used in scientific work.

A *thermometer* is an instrument for measuring temperature according to some definite scale. A *constant volume gas thermometer* is an apparatus for measuring temperature by the variation in pressure of a gas confined at constant or nearly constant volume. If the gas used is hydrogen the thermometer gives at once standard temperature; with other gases it must be calibrated in terms of the standard. Such an arrangement is shown diagramatically in Fig. 170, and consists essentially of a bulb of glass, glazed porcelain, fused quartz, platinum or platinum-iridium (according to the temperature range over which it is to be used), connected by a capillary tube to a mercury pressure-gauge such as the open manometer shown. The pressure of the confined gas can be measured by reading the difference in level of the two mercury columns and adding to this the atmospheric pressure as determined simultaneously with a barometer.

Still keeping the pressure of hydrogen at constant volume as the basis of the temperature scale, other numbers may be assigned to given temperatures by giving another number to the melting-point and subdividing the interval from melting to boiling into a different number of degrees. In this way the Fahrenheit scale (the one in ordinary use in English-speaking countries) is obtained by giving the value 32 to the freezing-point and subdividing the interval from the freezing-point to the boiling-point, the *fundamental interval* as it is called, into 180°. (However, Fahrenheit originally used other temperatures to define his scale, namely a freezing mixture of water, ice, and salt giving what he called 0°, and blood heat which he called 96°.) From the above statements we derive the following conversion formula for changing from one temperature scale to the other:

$$(t_F - 32)\frac{5}{9} = t_C$$

It must be clearly understood that the choice of a *thermometric property* (in this case *pressure* of *hydrogen*) is entirely independent

of the choice of numerical scale, *i.e.*, reference point and size of degree; the Centigrade or Fahrenheit numerical scale can each be applied to any other thermometric property desired.

It is found that the change in pressure (volume constant) of hydrogen for 1°C. as above defined is very closely $\frac{1}{273.0}$ of the pressure at 0°C.; hence if the same scale of temperature were carried below zero Centigrade (Fig. 171) the pressure would be reduced to zero at a temperature of about −273.0°C. This is called the *absolute zero of the hydrogen constant volume scale*, and, according to the ideas of the kinetic theory of gases (§227), it corresponds to a state of zero molecular velocity, since pressure is due to the impact of moving molecules. This temperature could not, however, be measured with the hydrogen thermometer, because, as we shall see, the gas would become liquid before this point was reached. We shall use T to represent absolute temperatures on the hydrogen scale. In order to give at once some idea of the known range of temperatures on the centigrade hydrogen scale it may be noted that:

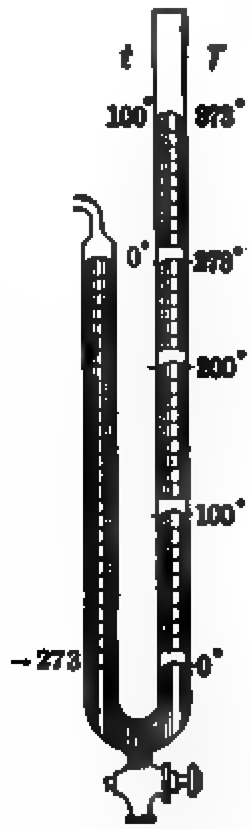


FIG. 171.—Temperature scale determined by change in pressure of a gas at constant volume. P_0 = 1 Atm. = external pressure.

−273.0° = absolute zero.
−270° = lowest temperature ever measured.
−190° = temperature of liquid air under 1 atmosphere pressure.
−80° = lowest recorded natural temperature.
0° = melting-point of ice.
100° = boiling-point of water under 1 atmosphere pressure.

700° = "dull red" heat for most solids.
1400° = "white heat" for most solids.
3800° = about the temperature of the electric arc.
6000°–7000° = Sun's temperature.

264. Constant Volume Gas Thermometer.—In order to use the constant volume gas thermometer in the simplest possible way to measure temperatures according to the standard hydrogen scale, it is evident that the volume of the bulb should be absolutely constant, and that all the gas used (including that in the capillary and over the mercury) should be heated to the temperatures to be measured. This is impracticable, and hence corrections must be made to the observed readings. Disregarding all corrections, we derive an approximate expression for the temperature of the bulb corresponding to a given pressure reading, as follows:
Let P_0 = pressure of hydrogen at the freezing-point of water,

P_{100} = pressure of hydrogen at the boiling-point of water,
t = some other temperature of the bulb, the value of which is to be determined, measured from centigrade zero.
P_t = pressure of hydrogen at this temperature t,

Then, in accordance with the definition of the degree (§263), we define any temperature t on the Centigrade scale of the constant volume hydrogen thermometer by the following formula:

$$t = \frac{P_t - P_0}{\frac{P_{100} - P_0}{100}} = \frac{P_t - P_0}{bP_0}$$

where b stands for the fraction $\frac{P_{100} - P_0}{100P_0}$, which, for hydrogen, is the same as $\frac{\text{increase in pressure for } 1°\text{C.}}{\text{pressure at } 0°\text{C.}}$
If now t_0 is such that $P_t = 0$, it follows that,

$$t_0 = -\frac{P_0}{\frac{P_{100} - P_0}{100}} = -\frac{1}{b},$$

That is to say, the absolute zero, at which $P = 0$, is $1/b$ degrees Centigrade below 0°C. Therefore letting T represent the value

of a temperature measured from absolute zero we have

$$T = t + \frac{1}{b} = t + T_0,$$

T_0 being the number on the absolute scale corresponding to 0° on the Centigrade scale.

The constant b is called the "coefficient of increase of pressure" or simply the *pressure coefficient;* for hydrogen its value is $\frac{1}{273.04}$, hence the value of the absolute zero of temperature on the centigrade constant volume hydrogen scale as defined above would be –273.04°. The value of b for air and nitrogen also is not very different from $\frac{1}{273}$, so that these two gases would give constant volume temperature scales approximately agreeing with the standard. Nevertheless it is obvious that the *exact* definition of the standard scale as here given is entirely dependent upon the properties of hydrogen. It has been found impossible, however, to use hydrogen above about 1100° C. because of the ease with which it passes through the walls of the metal bulbs which are best used at higher temperatures; under these conditions nitrogen is usually substituted.

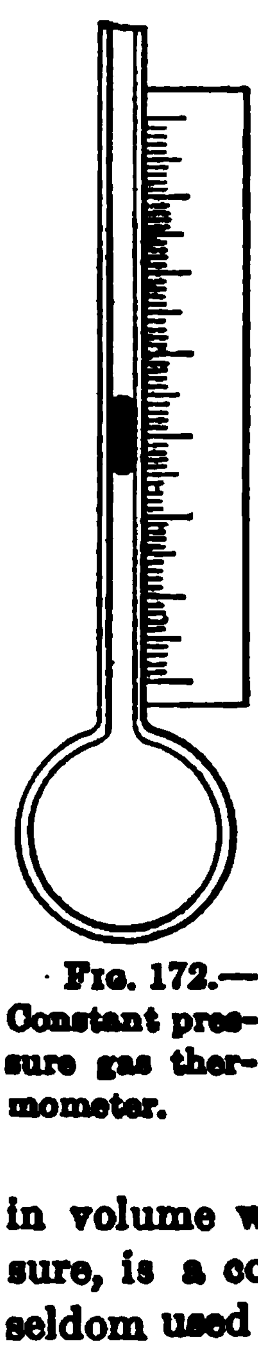


FIG. 172.—Constant pressure gas thermometer.

All gases increase in pressure if heated at constant volume, and the pressure at any temperature is given approximately by the equation

$$P_t = P_o(1 + bt)$$

where P_o is the pressure at 0°C. and b has somewhat different values for different gases (§279).

265. Constant Pressure Gas Thermometer.—The constant pressure gas thermometer, which makes use of the increase in volume with increasing temperature of a gas confined at constant pressure, is a convenient indicating device for demonstration purposes, though seldom used for precise measurements. As shown in Fig. 172, the *constant pressure* used is that of the external atmosphere, and the change in volume is proportional to the motion of an indicating globule of mercury or other liquid along a tube of uniform bore.

The *coefficient of expansion*, that is to say, the ratio $\frac{V_{100} - V_o}{100\, V_o}$, where V_o, V_{100} are the volumes at 0° and 100°C. respectively (pressure constant), is approximately $\frac{1}{273}$ for hydrogen, air, oxygen and nitrogen, so that an extremely sensitive indicator may easily be obtained.

With a bulb about 10 cm. in diameter and a tube 5 mm. in diameter the motion of the globule would be about 10 cm. per degree change in temperature of the bulb. The expansion of air when heated is one of the earliest known effects of heat, and the first thermometer, invented by Galileo in 1593, was based on this principle.

266. Mercury Thermometers.—For ordinary purposes thermometers depending on the expansion of mercury confined in a bulb and tube of glass or other transparent substance are most convenient and universally used. Two standard forms are shown in Fig. 173, the mercury being confined in a thin-walled glass bulb attached to an extremely fine capillary tube. For use at ordinary temperatures the upper part of the capillary contains only mercury vapor. Since mercury expands somewhat less than $\frac{1}{5000}$ part of its volume at 0°C. for a degree rise in temperature (compare with air above), it is necessary to have a very fine capillary in order to obtain an easily observable motion of the column for a degree change in temperature. All such thermometers should, for precise work, be calibrated or standardized by comparison with the hydrogen standard.

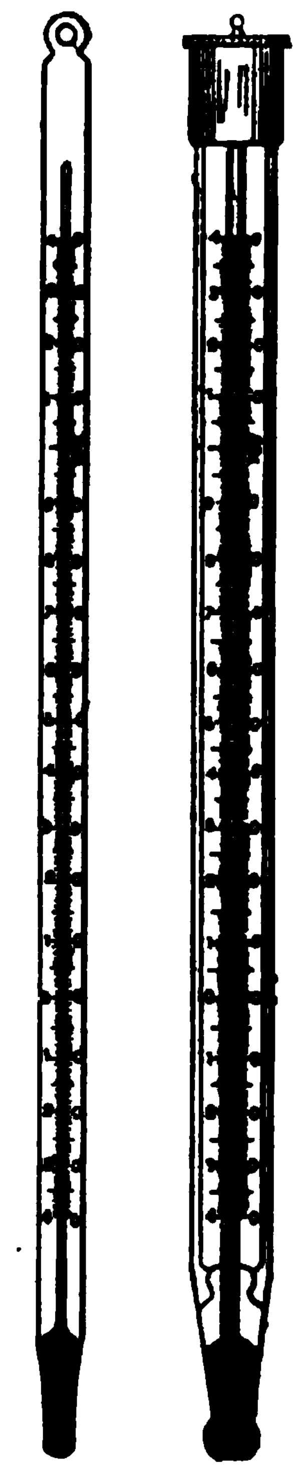


FIG. 173.—Thermometers with solid stem and with enclosed scale.

Fig. 173 shows the two standard ways of marking the "scale" on the thermometer. In one the scale is marked directly on the stem of the thermometer—this is the most accurate and permanent way, used in all standard scientific thermometers and clinical thermometers; in the other the scale is on paper or white glass and enclosed in an outer glass tube back of the capillary stem—this usually gives more legible scales but they are somewhat likely to become loose and shift with respect to the capillary. A third method is used for cheap "household" thermometers; in this the thermometer is simply mounted on a support which carries the scale.

The glass used for the thermometer (especially the *bulb*) is of the greatest importance, and in recent years great improvements have been made in the qualities of glass used for this purpose. A bulb made of ordinary glass has the fault of slowly changing its volume with time, and of permanently and quickly increasing its volume whenever it is heated, say to 100°C. or higher. Such changes, of course, alter the reading for a given temperature. Some of these effects gradually disappear after the bulb has been made; so that

thermometer bulbs should be kept for some time, or else artificially "aged" by heating and cooling, before being graduated.

Through the development of special glasses having high melting-points it has become possible to construct mercury-in-glass thermometers reading to 550°C. or even higher. In such high-range thermometers the space above the mercury column must be filled with a gas (usually carbon dioxide or nitrogen) at a final pressure of about 19 atmospheres, in order to keep the mercury from boiling. For such thermometers the properties of the glass are of the greatest importance, and the glass known as "Jena 59^{III}" is the best one to use. Even with this glass if the thermometer is kept at 550°C. for an hour or more a permanent expansion of the bulb will result: this will permanently lower the freezing-point reading, but if this change is applied as a correction (added) to subsequent readings of the thermometer, fairly

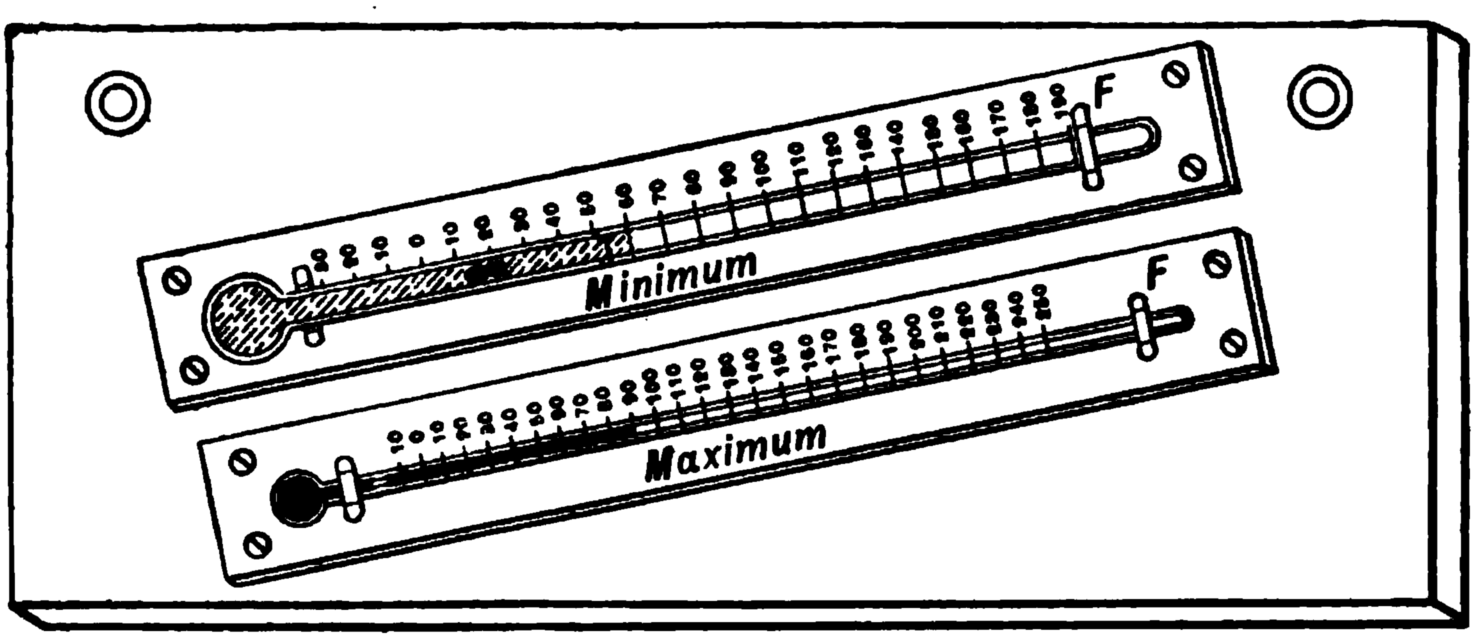


FIG. 174.—Maximum and minimum thermometers.

correct results can be obtained. Thermometers of mercury in clear fused quartz have also recently been satisfactorily constructed for use up to about 700°C.

In using thermometers it is well always to avoid too sudden heating or cooling; and in measurements above 100° (or in all cases where extreme accuracy is required) it must be remembered that thermometers are usually graduated to read correctly when *bulb and stem* are all at the temperature to be measured. If the stem is cooler than the bulb the thermometer will read too low and this error may amount to as much as 40° at 550°C. In careful work thermometers should always be compared with a standard, or standardized at known temperatures (§271) or sent to the Bureau of Standards for comparison.

267. Special Forms of Thermometers.—Alcohol and some other liquids have an advantage over mercury in their greater coefficient of expansion and smaller surface tension (giving more

regular rise and fall in the capillary), but they are seldom used for accurate thermometers. Since mercury freezes at $-38.8°C.$, thermometers containing alcohol are often used for temperatures below this. Pentane (C_5H_{12}) is also used for thermometers reading to $-190°C.$

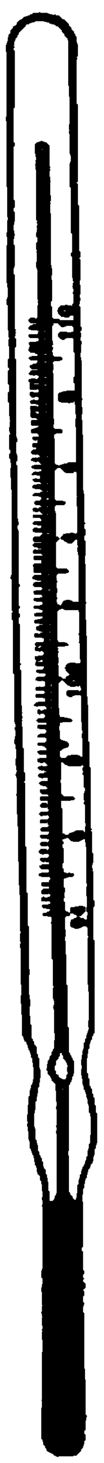


FIG. 175.—Clinical thermometer.

Maximum and *minimum* thermometers are thermometers provided with devices for recording the maximum or minimum point reached by the end of the mercury column. The maximum thermometer is usually of one of two forms. In the first form, a small iron index is pushed ahead of the mercury column and left when the column contracts, the *lower* end of the index indicating the highest reading of the mercury column; in the second form, Fig. 174, a contraction is made in the bore of the tube near the bulb and at this point the mercury column *breaks*, when contraction occurs after the maximum point is reached, leaving the upper end of the column at the maximum reading. This device of a contracted bore is used in clinical thermometers, Fig. 175. Minimum thermometers, Fig. 174, are usually of alcohol in glass, and have below the meniscus a light index, of such form that the alcohol can flow past it, while it will be dragged *down* when the descending meniscus reaches it. If the thermometer is kept nearly horizontal, the index will rest at the *lowest* point reached by the meniscus.

For some purposes (especially common thermostats) metallic thermometers are used. They usually depend upon the bending of a duplex metallic bar, Fig. 181, because of the different amounts of expansion of its component metals. They are not saistfactory for accurate work.

268. Resistance Thermometry.—In recent years an electrical method of thermometry has come into very general use. In this the thermometric property is the resistance offered by a metallic wire to the passage of an electric current, which resistance changes with the temperature. It must be remembered that such thermometers like all secondary instru-

14

ments, must be *calibrated in terms of the hydrogen standard.* On account of its permanence, high melting-point and acid-resisting qualities, pure platinum wire has been most extensively used for this purpose, though for use at ordinary temperatures copper and iron wire may be substituted. The usual form of platinum-resistance thermometer is shown in Fig. 176, the coil whose resistance changes are to be measured (called by analogy the "bulb" of the thermometer), being mounted in a protecting tube of glass, or (which is better) of metal for moderate temperatures and of porcelain for high temperatures.

The advantages of the platinum thermometer are permanence and reliability, wide range (it may be used up to 1200°C.), the fact that the readings may be made at a distance of several hundred feet from the thermometer

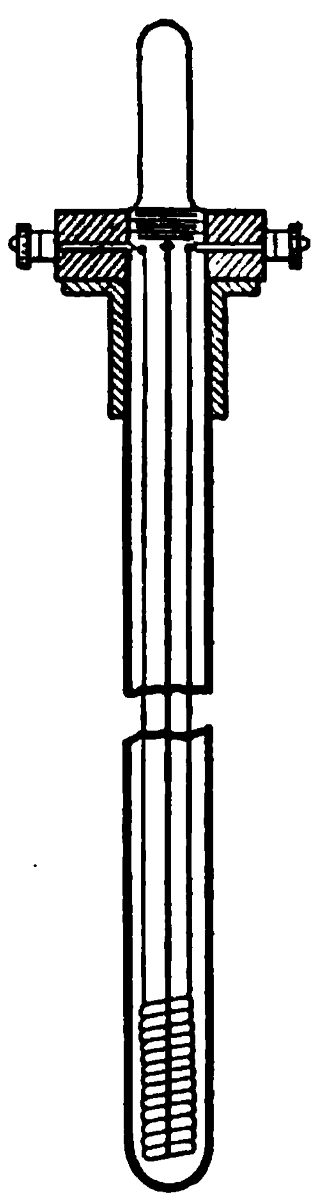


FIG. 176.—Platinum resistance thermometer.

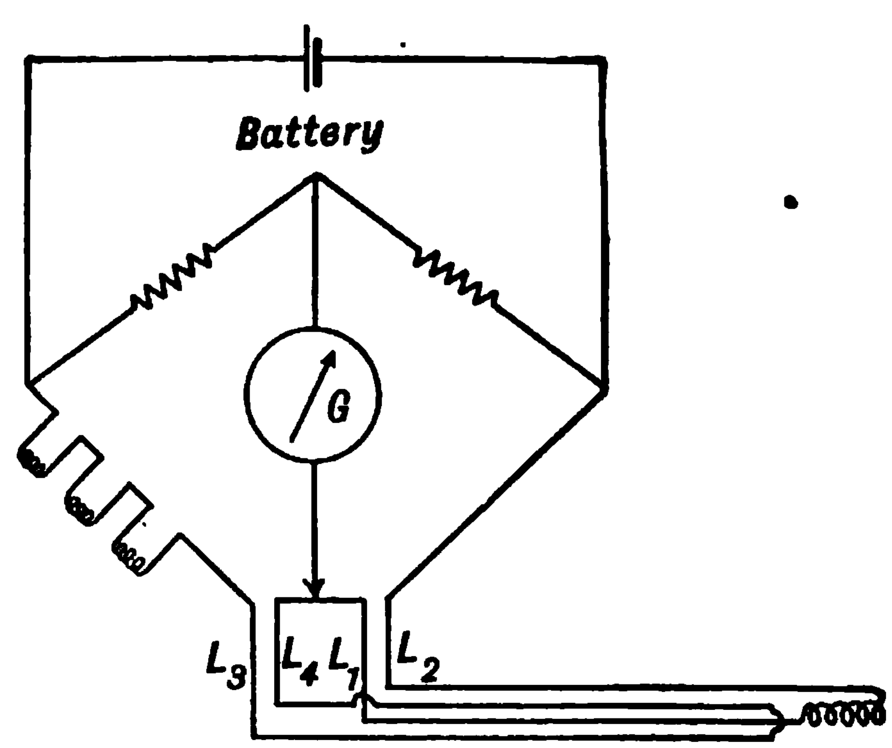


FIG. 177.—Wheatstone's bridge for measuring resistance of platinum thermometer.

itself and that it may be made accurately self recording. It is also capable of extreme sensitiveness, $\frac{1}{10000}$°C. being readable. For all these reasons its use in scientific and engineering work is rapidly increasing.

Fig. 177 shows the electrical leads to the coil L_1 L_2, compensating leads L_3 L_4 by means of which the effect of temperature changes in the leads L_1 L_2 are eliminated, and the connection of the Wheatstone bridge (see §456) by which the resistance is measured. From an empirical formula developed

by Callendar the temperature corresponding to a given resistance may easily be obtained. This formula is of such a form that only three known temperatures are needed to determine its constants. It is, therefore, very easy to standardize a platinum thermometer.

269. Thermo-electric Thermometer.—When two different metals are joined together in a circuit as shown in Fig. 178, and one *junction* is heated, an electromotive force is in general produced (see §477), which tends to drive a current in a certain direction as shown and this electromotive force increases as the difference in temperature between the two junctions increases. This thermal electromotive force is another thermometric property very extensively used. For some purposes a voltmeter (see § 453) suffices to measure the electromotive force generated by heating one junction, and it may be calibrated to read temperature directly. The thermoelectric thermometer or thermo-couple, as it is called, is valuable on account of its sensibility, quick response to temperature changes, and the small size and mass of the part which must be heated as compared to the bulb of a mercury or resistance thermometer.

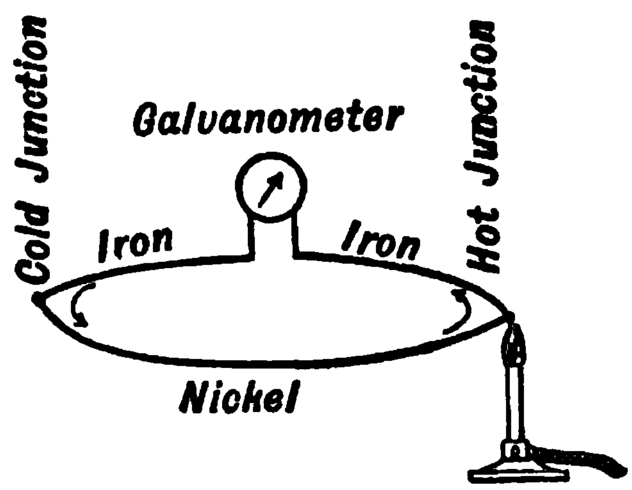


FIG. 178.—Thermoelectric couple, showing direction of current produced by heating.

For work below 500°C., wires of copper and constantan (an alloy of copper and nickel) are quite satisfactory; up to 1000°C. wires of nickel and nickel-chromium alloy may be used for approximate work, while, for the entire range up to 1600°C., the most accurate results are given by wires of platinum and platinum +10% rhodium.

270. Measurement of High and Low Temperatures.—The measurement of extreme high or low temperatures presents separate and difficult problems. This is partly because of mechanical difficulties caused by changes in properties of ordinary substances at extreme temperatures (for example, melting and softening of metals and porcelain), and chemical reactions at high temperatures, and partly because the range of the direct hydrogen thermometer is passed and it is necessary to *extrapolate* by means of some empirical formula. At high temperatures recourse is had to nitrogen in a constant volume thermometer, which has been used from 1100°C. to 1550°C.; above this for a short range thermoelectric extrapolation is possible, while beyond this a radiation scale (see §339) and radiation methods are the only resources. At low temperatures the least liquefiable gas, *helium*, used in a

constant volume thermometer, but at a pressure of only 10 cm. of mercury, has been used, as well as the resistance and thermoelectric methods.

271. Standard Temperatures.—For the purpose of standardising thermometers, thermo-couples, resistance thermometers, etc., it is convenient to make use of one or more temperatures which can easily be obtained and kept constant, and which have been accurately measured. For such purposes melting- and boiling-points are the most convenient. To use a standard boiling-point the liquid must be steadily boiled at a known pressure and the thermometer immersed in the *vapor;* to use a melting-point the thermometer may be immersed in a mixture of the solid and liquid. The following table gives some of the more useful points.

TABLE 1

STANDARD TEMPERATURES

(Pressure Constant at One Atmosphere)

Hydrogen (liquid)	Boiling-point,	−253°C.
Oxygen	Boiling-point,	−183
Carbon dioxide	Boiling-point,	− 78.2
Mercury	Melting-point,	− 38.8
Water	Melting-point,	0
Ether	Boiling-point,	34.6
Alcohol (ethyl)	Boiling-point,	78.3
Water	Boiling-point,	100
Napthalene	Boiling-point,	218.0
Tin	Melting-point,	231.9
Benzophenone	Boiling-point,	306.0
Sulphur	Boiling-point,	444.7
Sodium chloride	Melting-point,	801
Silver	Melting-point,	960
Gold	Melting-point,	1063
Palladium	Melting-point,	1549
Platinum	Melting-point,	1753

272. The Pressure, Volume, Temperature Diagram.—From the discussion of §262 we saw that in order to know the *condition* of a body we should know the amount of energy present, per unit mass, in several different forms, namely as kinetic energy of molecules, atoms and electrons and as potential energy of molecules, atoms and electrons. Of the entire amount of this internal energy we have no knowledge, but we can measure the heat energy which passes into or out of a substance and also the external work done, which together constitute the *change* in the internal energy, and hence we can tell when a body is brought back to a given condition of total internal energy. Now it is found that in the majority of cases when a body is brought back

to the same total energy content its pressure, volume and temperature return to the same values, and, in fact, all its physical properties are the same as before; hence it is said that the pressure, volume, and temperature *determine the physical state of a body.*

These three variables, P, V, and t are, however, not independent but are connected by a relation, called an *equation of state*, the general form of which is not known. This relation expresses the experimental fact that if we fix any two of the three variables, P, V, and t, the third must have a definite value. For example, if a gas occupies a given volume at a given pressure it must have a certain temperature.

Since the physical condition of a body is determined by the values of the three variables, P, V, and t, it is very natural to represent a given condition by a point having the corresponding values of P, V, and t as coördinates measured along three rectangular axes, as in Fig. 179, where every point in space represents a definite physical condition. If we take as the origin *absolute zero* values of P, V, and t, then negative values of V and t will mean nothing physically, while negative values of P will mean tensions. Points in a plane parallel to the PV plane will correspond to physical conditions for all of which the temperature is constant, and, similarly, planes parallel to the tV and Pt planes respectively will represent constant pressure and constant volume conditions. Since it is usually sufficient to fix two of the variables P, V, and t, physical conditions are often represented by points in a plane, for which purpose the PV, Pt, or Vt plane may be chosen.

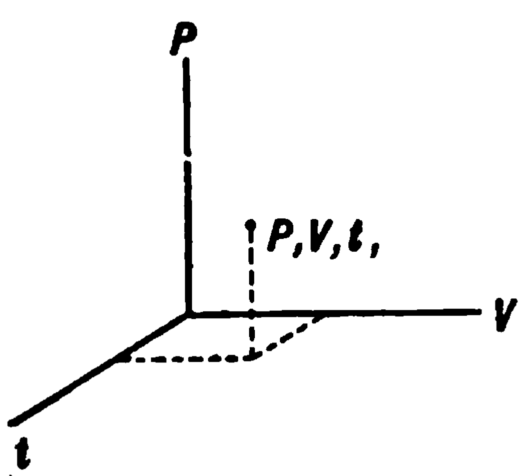


FIG. 179.—The use of P, V, t as coördinates.

EXPANSION

273. Introduction.—The important changes in substances produced by heat are changes in size, changes in the arrangement of molecules with respect to one another, and changes in state, from solid to liquid and gaseous. The difference between solids, liquids and gases has been discussed in §157. Solids in general offer great resistance to change of shape, and their molecules

tend to assume a definite arrangement in groups called crystalline structure, not only in obviously crystalline minerals such as quartz, but in all solids. The existence of such structure is sometimes taken as a test for the solid state, though liquids also can have crystalline properties, and it is difficult to draw a sharp distinction between the two. From the heat standpoint the important matters are that the average molecule in a solid moves about much less than in a liquid or gas, and that the potential energy of the molecules with respect to each other is greatest in the gaseous state; furthermore the potential energy of a solid, liquid or gas changes with its change of size, or expansion due to heat. In discussing the expansion of solids it is convenient to consider both their change in linear dimensions and their change in volume, while for fluids the latter alone has a meaning.

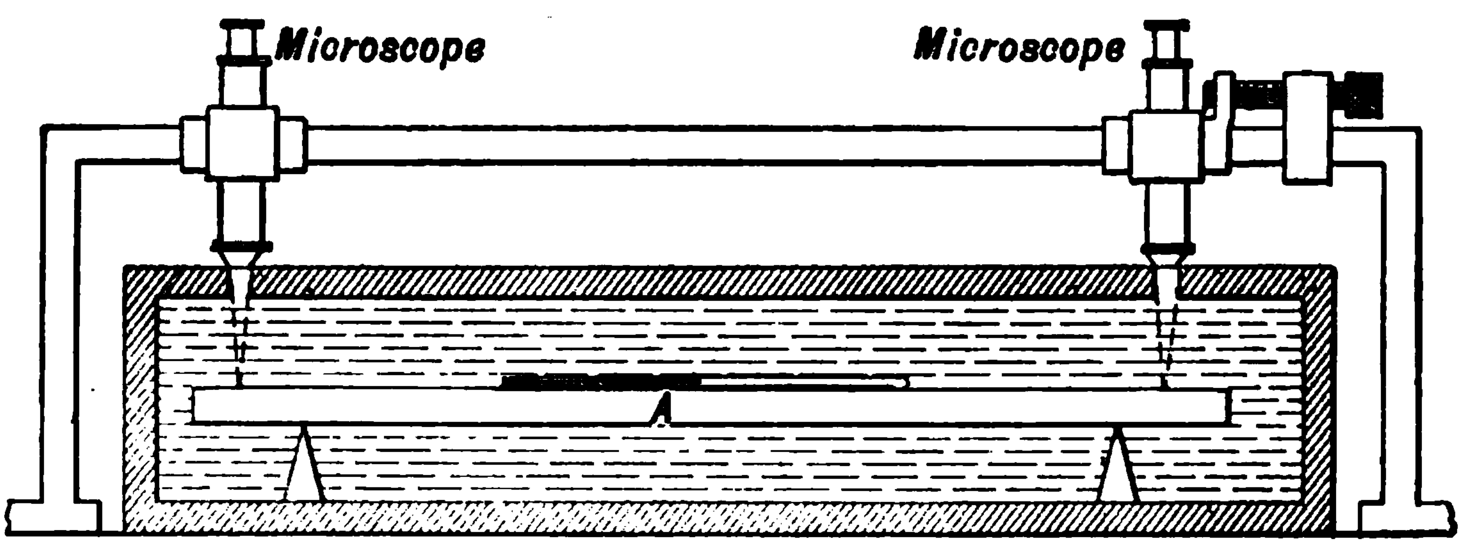


FIG. 180.—Apparatus for measuring coefficient of linear expansion of solids.

274. Linear Expansion of Solids.—This is an effect very easily observed and very widely made use of. Telegraph wires which sag in summer are taut in winter; the tires of wagon and locomotive wheels and jackets of large cannon are made too small to slip in place and are then put on while expanded by heat, so that when cool and shrunk they have a firm grip. Different solids expand differently for the same change in temperature. A simple experimental arrangement for measuring the amount of expansion is shown in Fig. 180 where A, a bar of the material being studied, is supported in a bath so that its temperature may be varied. Two microscopes, supported by a frame distinct from the bath, are arranged so that one or both may be moved parallel to the bar by a fine micrometer screw and focused on two fine marks

made on the bar. As the bar expands the microscopes are moved so that the cross-hairs remain set on the marks, and thus the expansion can be read from the graduated heads of the micrometer screws. By substituting a standard meter for the bar, the actual length between the marks at any desired temperature, say 0°C., may be determined, and, by adding to this the observed expansions, the length L_t of the bar at any temperature t may be obtained. The expansion will usually be found to be approximately, though not exactly, proportional to the change in temperature, that is to say, if the values L_t are plotted as ordinates with the corresponding values of t as abscissæ, the result will be a curve, though the curvature is usually slight. In general it is found that L_t may be very closely represented by an expression of this form,

$$L_t = L_0(1 + at + bt^2 + ct^3 + \cdots) \qquad (1)$$

where a, b, c are constants and t is the temperature on the Centigrade scale. The number of constants necessary increases with the temperature range over which it is attempted to work and with the accuracy desired, and varies also with different substances. For small temperature differences a, usually called "the coefficient of expansion" is sufficient, and its value is evidently

$$a = \frac{L_t - L_0}{L_0 t}$$

Frequently also a *mean coefficient of expansion* between two temperatures, t_1 and t_2, is used, and its value is accordingly

$$a_m = \frac{L_2 - L_1}{L_1(t_2 - t_1)}$$

For moderate ranges of temperature (*e.g.*, 0° to 100°) a and a_m usually differ so little that they need not be distinguished.

As may be seen from the following table, the coefficients of expansion are never large, and very refined experimental methods are necessary to determine them accurately, as, for instance, some form of *interferometer* (§722).

TABLE 2

COEFFICIENTS OF LINEAR EXPANSION

Substance.	Cm. per degree C. per cm.
Aluminum	25.5×10^{-6}
Brass	18.9 "
Copper	16.7 "
Glass (Jena 16^{III})	7.8 "
Gold	13.9 "
Hard rubber	80. "
Ice	50.7 "
Invar	0.7 "
Iron (cast)	10.2 "
Iron (wrought)	11.9 "
Lead	27.6 "
Nickel	12.8 "
Oak, ‖ grain	4.9 "
Oak, ⊥ grain	54.4 "
Platinum	8.9 "
Porcelain (Berlin)	2.8 "
Quartz, ‖ axis	7.5 "
Quartz, ⊥ axis	13.7 "
Quartz, fused	0.39 "
Silver	18.8 "
Tin	21.4 "
Zinc	26.3 "

Isotropic solids, including crystals in the cubical system (with three equal axes of symmetry), expand equally in all directions. Other crystals have one axis of symmetry, with one coefficient of expansion along this axis and another one in a plane at right angles to the axis, the coefficient being the same in all directions in this plane; while still others have three different expansions along three axes, in some cases even showing a contraction along one axis. In such cases of unequal expansions the angles of a crystal change as the crystal expands.

275. Applications of Linear Expansion.—The expansion of solids, especially the differential expansion, is made use of in metallic thermometers, thermographs and thermostats. Usually a compound strip of brass and iron, riveted together, is fixed at one end and arranged so that the bending of the strip, due to the unequal expansion of brass and iron, operates a recording or indicating pointer, or, in the thermostat, makes electrical

contact to right or left and thus controls some heating system Fig. 181 shows a common form.

The balance wheel of watches has a rim made of a compound metal strip, as above described, and so arranged that a change in temperature, by altering the curvature of these strips, will move a considerable part of the mass of the wheel to or from the center, thus altering the moment of inertia of the wheel and hence its period. In this way other temperature effects on the rate of the watch, such as change in elasticity of the springs, change in diameter of the balance wheel, and change in viscosity of the oil in the bearings, may be compensated.

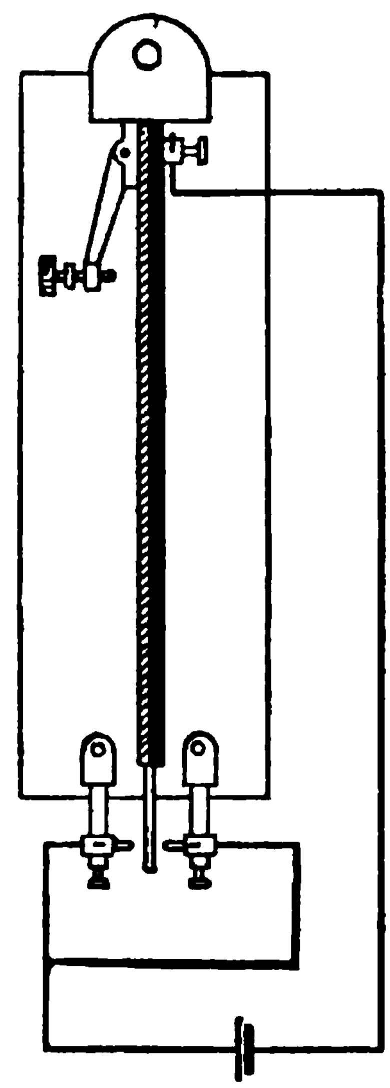


Fig. 181.—Metallic thermometer depending upon difference in expansion of two metal strips, arranged as a thermostat.

In the mercury clock pendulum shown in Fig. 182 the length of the reservoir of mercury is so chosen that the expansion of the mercury, which raises the center of gravity, just compensates for the expansion of the supporting rod which lowers the center of gravity, so that the time of vibration will not be altered by changes in temperature. This compensation can now be accomplished even more accurately by the use of a specially worked nickel-steel alloy, called "invar," which has a coefficient of only .00000075 to .00000015, or $\frac{1}{10}$ that of brass. This alloy is also valuable for making standard meter-bars, tapes and scales with lengths practically independent of small temperature variations.

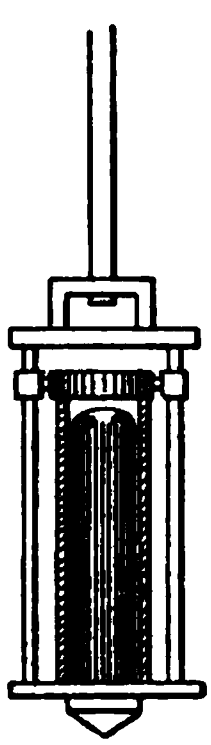


Fig. 182.—Mercury compensation pendulum.

The cracking of objects by heating, particularly sudden heating, is due to unequal expansion produced by differences of temperature in different parts. Porcelain is less liable to crack than glass because of its smaller coefficient of expansion, and thin glass than thick because of the more rapid equalization of temperature. In fusing metals into glass to make an air-tight joint, as, for example, the leads into an incandescent

lamp bulb, it is necessary to use a metal having nearly the same coefficient of expansion as glass, otherwise cracking (or leaking) would occur when the joint cooled. As may be seen from the table, platinum is the best metal for this purpose. There is a very striking difference between the coefficients of crystalline and fused quartz; the former cracks with the slightest heating, the latter, because of its small coefficient, may be taken from an oxyhydrogen flame and at once plunged into liquid air without cracking.

276. Cubical Expansion of Solids.—If V_t represents the volume of a solid at t°C., V_0 its volume at 0°C.; then it is found that in general solids expand in such a way that V may be represented as a function of t by an equation similar to the one used for linear expansion:

$$V_t = V_0(1 + a't + b't^2 + c't^3) \qquad (1)$$

and, as before, the constants, b', c', etc., are much smaller than a', so that for small temperature changes,

$$V_t = V_0(1 + a't) \qquad (2)$$

If we now consider a cube of the material of length L_t on an edge, we have,

$$V_t = L_t^3 = L_0^3(1 + at)^3,$$

approximately, from equation (1), § 274, or

$$V_t = L_0^3(1 + 3at),$$

neglecting higher powers of a; and, since $V_0 = L_0^3$,

$$3a = a', \text{ by comparison with equation (2)}$$

That is to say, *the cubical coefficient of expansion is three times the linear coefficient*, and can be obtained from the preceding table.

277. Expansion of Liquids.—The change of volume of liquids with temperature has already been mentioned as the basis of liquid-in-glass thermometers. The fact that the mercury or alcohol in such thermometers rises with increased temperature shows that the liquids expand more than the glass, and this is usually true of liquids as compared with solids. To represent the volume V_t of a liquid at a temperature t in terms of the volume V_0 at 0°C., it is found that an equation of the same form will suffice—

$$V_t = V_0(1 + a''t + b''t^2 + \cdots)$$

or approximately,

$$V_t = V_o(1 + a''t)$$

since b'' is usually much smaller than a''.

A bulb with a capillary stem like a thermometer is usually used in measuring the *differential expansion* of a liquid and a solid. The walls of the bulb expand as if they were filled with solid material; hence the volume of the bulb space is at any temperature equal to the expanded volume of the solid which would fill it at 0°C. If

V_o = volume of bulb and of liquid filling bulb at 0°C.,
V'_t = volume of bulb at t°C.,
V_t = volume of same liquid at t°C.,
a', a'' = volume coefficients of expansion of solid composing the bulb, and of the liquid respectively, then

$$V_t - V'_t = V_o[(1 + a''t) - (1 + a't)] = V_o(a'' - a')t.$$

This *differential or apparent expansion*, $V_t - V'_t$, can be measured by noting the rise of the liquid in the capillary stem. If, in addition, the volume coefficient a' of the solid is known, we can determine the coefficient a'' of the liquid; for—

$$a'' = \frac{V_t - V'_t}{V_o t} + a'$$

The difference $a'' - a'$ is called the *apparent coefficient of expansion* of the liquid.

It is possible to determine the *absolute coefficient of expansion* of a liquid, independent of the expansion of a containing vessel, by a method due to Dulong and Petit and illustrated in its simplest form in Fig. 183. Two vertical tubes filled with the liquid in question are connected at their lower extremities by an accurately horizontal tube. The vertical tubes are in baths of some sort, so that one can be maintained at a temperature of 0°C. and the other at t°C. At the bases of the two tubes the pressures must be equal, otherwise there would be a flow from one to the other through the connecting tube. The pressure at the two upper free surfaces must be the same, since it is that of the external atmosphere; hence the difference in

Fig. 183.—Method of measuring the absolute coefficient of expansion of mercury.

pressure from top to bottom of the two columns must be the same. Hence by §185,

$$h_t \rho_t g = h_o \rho_o g$$

and $$\frac{\rho_o}{\rho_t} = \frac{h_t}{h_o}$$

But if V_o = volume of unit mass of fluid at 0°C. $= \frac{1}{\rho_o}$

and V_t = volume of unit mass of fluid at t°C. $= \frac{1}{\rho_t}$

then $V_t = V_o(1 + a''t)$

and $$\frac{\rho_o}{\rho_t} = \frac{V_t}{V_o} = 1 + a''t = \frac{h_t}{h_o}$$

Hence $$a'' = \left(\frac{h_t}{h_o} - 1\right)\frac{1}{t}$$

This method has been especially used to determine the absolute coefficient of expansion of mercury; this being known, mercury can be used to determine the coefficient of expansion of solids by the differential method. The coefficients of expansion of liquids (except water) decrease with increase of the pressure at which they are observed.

TABLE 3

COEFFICIENTS OF CUBICAL EXPANSIONS OF LIQUIDS

Substance.	Cm³. per degree C. per cm³.	
Alcohol (ethyl)	110.	$\times 10^{-5}$
Alcohol (methyl)	118.	"
Benzine	124.	"
Mercury	18.18	"
Paraffin oil	90.	"
Pentane	159.	"
Toluene	109.	"
Water, 15–100°	37.2	"
Xylol	101.	"

278. Expansion of Water.—Water is unique among liquids in that it has a maximum density at about 4°C., under 1 atmosphere pressure, *i.e.*, below 4°C. it contracts with rise of temperature, above 4°C. it expands.

This property, which has very important consequences, is clearly shown by Hope's apparatus, Fig. 184. If the tank around the middle of the glass vessel be filled with a freezing mixture of ice and salt, and the vessel be filled with water at a temperature

higher than 4° C., the water in the middle when cooled will become denser and fall to the bottom, whereas the water above the middle will not be disturbed. Thus the upper thermometer will indicate a practically stationary temperature and the lower one a falling temperature until all the lower half of the vessel is filled with water at 4°C., after which the upper one begins to fall in temperature until 0° is reached and freezing begins at the top, the lower thermometer still indicating 4°C. The water at 4°C. is most dense and therefore collects at the bottom of the vessel.

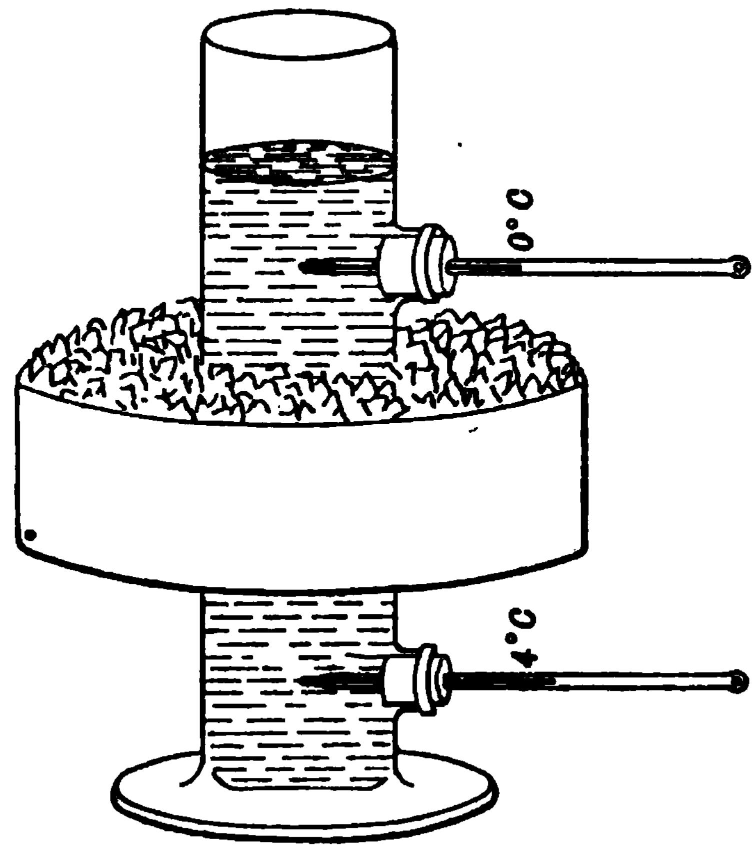


FIG. 184—Hope's apparatus for determining the temperature of maximum density of water.

A somewhat similar operation goes on in winter in ponds and rivers which are not too much disturbed by winds or currents, the densest water, at 4°C., collects at the bottom, while the coldest, at 0°C., being lighter, stays on top. Hence freezing occurs at the top, unless the entire mass of water is cooled by currents to near 0°C., in which case freezing may occur on the bottom or on submerged solids, cooled by radiation, thus forming *"ground ice"* which is of serious consequence in northern rivers. The volume of 1 gram of water at various temperatures under 1 atmosphere pressure is given in Fig. 185.

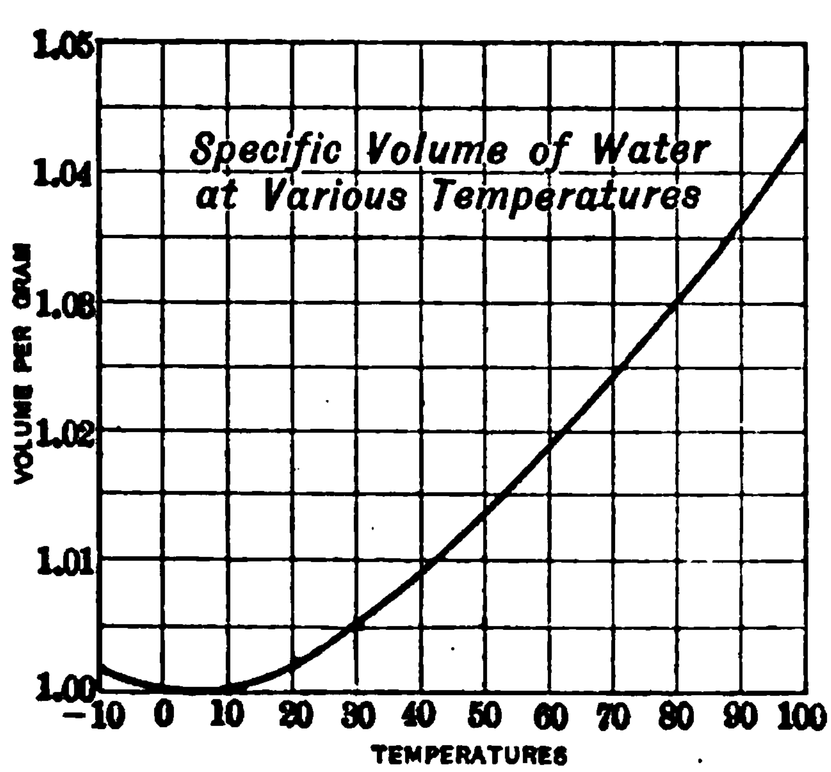


FIG. 185.—Expansion curve for water.

According to Amagat the temperature of maximum density falls with increase of pressure, being about 2°C. under a pressure of 93 atmospheres. If

kept by pressure in the liquid state water continues to contract below 0°C. and to expand at an increasing rate above 100°C. The solution of various salts in water also lowers the temperature of maximum density, 4 per cent. of dissolved common salt lowering it to -5.63°C. The peculiar behavior of water as regards its thermal expansion is due, according to Tamman, to the existence at low temperatures of several different kinds of water molecules or groups of molecules which gradually break up into one simpler kind as the temperature is raised.

279. Expansion of Gases.—Since the effect of pressure on the volume of a gas is very great (§221), it is evident that in discussing the expansion of gases with increase in temperature we must be careful to specify the pressure conditions which are to hold during the expansion. The simplest condition is to maintain the *pressure constant* and measure the change in volume of the gas in a bulb by allowing it to expand and push out a mercury piston in an attached tube, as illustrated in Fig. 172. For accurate work an arrangement such as is shown in Fig. 186 is necessary, and for more complete knowledge of the subject the expansion must be carried out at various constant pressures. A correction must, of course, be made for the expansion of the bulb and for the fact that an increasing amount of the gas will be in the stem and hence will not be heated. Gay-Lussac (1802) and Charles (1787) independently carried out such experiments, and arrived at the "Law of Charles and Gay-Lussac," according to which *all the common gases expand by a constant fraction of their volume at 0° for each rise of 1° in temperature.* This fraction is about .003660 ($\frac{1}{273}$), or about the same as the "pressure coefficient" of a gas (§264). In the form of an equation this law is—

FIG. 186.—Apparatus for measuring the expansion of gases. The gas expanding from the bulb in steam is measured in the graduated bulb.

$$V_t = V_o(1 + at) \qquad (p \text{ constant}),$$

but it is now known that the law is only approximately true and that a is not the same for all gases. Furthermore, a varies with

the pressure and with the temperature, and is not, in general, quite equal to the pressure coefficient, b.

Later work of Regnault and others has shown that the coefficients of expansion of all gases except hydrogen increase, at ordinary temperatures, with increasing density of the gas, and that the coefficients for the several gases are more nearly alike and more nearly equal to their "pressure coefficients" when the gases are at low pressures or high temperatures.

TABLE 4

EXPANSION COEFFICIENTS AND PRESSURE COEFFICIENTS

Gas.	a.	b.
Air	0.003671	0.003674
Carbon dioxide	0.003728	0.003712
Hydrogen	0.003661	0.003662
Nitrogen	0.003673	0.003672

Temperature 0°C. – 100°C., Pressure, 1 atmosphere.

280. The Gas Equation: A Perfect Gas.—We have seen (§221) that gases follow Boyle's law more or less closely, the product of the pressure and volume at constant temperature being nearly constant. In §264 we considered the change in pressure with temperature of a gas confined at constant volume, which is given approximately by the equation $P_t = P_o(1+bt)$. In §279 we have just discussed the expansion of gases, which occurs approximately according to the relation $V_t = V_o(1+at)$, the pressure being constant. It is convenient to combine these statements into a single equation, which will then represent all the relations which approximately hold between the pressure, volume and temperature of a gas. This may be done as follows:

Let P_oV_o be the pressure and volume of a given mass of gas at 0°C., and let it be heated at constant volume to t°C. Then we have (§264)

$$P_t = P_o(1+bt)$$

and hence

$$P_tV_o = P_oV_o(1+bt)$$

Again, starting at $P_oV_o0°$, let it be heated at constant pressure to the same final temperature t°C.; then by the law of Charles (§279)

$$V_t = V_o(1+at)$$

and hence

$$P_oV_t = P_oV_o(1+at)$$

Now let the pressure, P, and the volume, V, be changed in any way, the temperature remaining t°C. Then from Boyle's law

$$(PV)_t = P_o V_t = P_t V_o$$

Hence

$$(PV)_t = P_o V_o (1 + bt)$$

and

$$a = b$$

Also since (§264) $$t + \frac{1}{b} = T$$ bt + 1 = bT

We may write $$PV = P_o V_o b T$$

or $$PV = RT \qquad (5)$$

It is frequently convenient to consider an imaginary *ideal* or *perfect* gas, which exactly obeys these gas laws, and which also has certain other properties which will be referred to later. The volume and pressure coefficients of such a gas we shall designate by a' and b', absolute temperature according to this *perfect gas* scale by T', and the constant factor by R'. The gas law or equation of state for a perfect gas then becomes $PV = R'T'$.

281. Real Gases.—As we have just seen, real gases follow more or less closely the law

$$PV = RT$$

where T is the temperature measured with a constant volume hydrogen thermometer from the absolute zero of the hydrogen scale. The approximation to the law $PV = RT$ is found to be very much closer for high temperatures and low pressures.

The value of R is different for different gases, and, of course, also for different masses of any one gas. It is customary to consider the equation as applying to 1 gram of gas, and R is then called the *gas constant* for this gas, and is evidently equal to $\frac{1}{273.0}$ of the product $P_o V_o$ at 0°C. For any other mass of M grams the constant in the equation $PV = RT$ will be MR, since volumes are proportional to masses under given conditions.

It is easy to see, in a general way, why the properties of real gases should approach those of a perfect gas at high temperatures and low pressures. For, according to the simple kinetic theory (§227), a gas having no molecular forces, *i.e.*, *no molecular potential energy*, and *negligible molecular volume*, is perfect in so far that it obeys the law $PV = RT$. Now, it is evident that the

higher the temperature of a real gas, the less will be the proportion of the potential to the kinetic energy, and also that the larger the volume of a gas, other things being equal, the less will be the actual molecular volume compared to the total volume. Hence, as the temperature is raised, or the density diminished, the conditions become more nearly those assumed in the simple kinetic theory.

By making a still further assumption equation (5) may be further generalized. According to Avogadro's hypothesis equal volumes of different gases at the same temperature and pressure contain equal numbers of molecules, that is, the total masses of equal volumes will be proportional to the molecular weights of the gases, or

$$M_1 : M_2 : M_3 \ . \ . \ . = m_1 : m_2 : m_3 \ . \ . \ .$$

where m_1, m_2, m_3 are molecular weights. Hence, if we take m_1 grams, m_2 grams, and m_3 grams (called gram-molecular-weights) of these gases, they will occupy the same volume at the same pressure and temperature.

Hence, $$PV = m_1R_1T = m_2R_2T = m_3R_3T$$

Hence, $$m_1R_1 = m_2R_2 = m_3R_3 = R''$$

and R'' is a constant for all gases, whose value can be at once computed. For example, for nitrogen $m = 28$; specific volume $V = 796.2$ c.c. when $T = 273°$; and $P = 1$ atm. $= 1{,}012{,}630$ dynes/cm.^{2}
Hence

$$R'' = \frac{PVm}{T} = 8.305 \times 10^7 \ \frac{\text{ergs}}{\text{degree}}$$

282. Isothermal Curves.—The significance of the equation $PV = RT$ can be seen more readily by graphical representation according to the method of §272. Giving T some constant value, T_1, it is evident that Boyle's law, PV = const., is represented by a rectangular hyperbola in a plane parallel to the PV plane and cutting the T axis in the point T_1. If a series of such hyperbolæ are located for different temperatures and then *projected* upon the PV plane by dropping perpendiculars from every point to this plane, the result is a *family* of hyperbolæ each of which can be

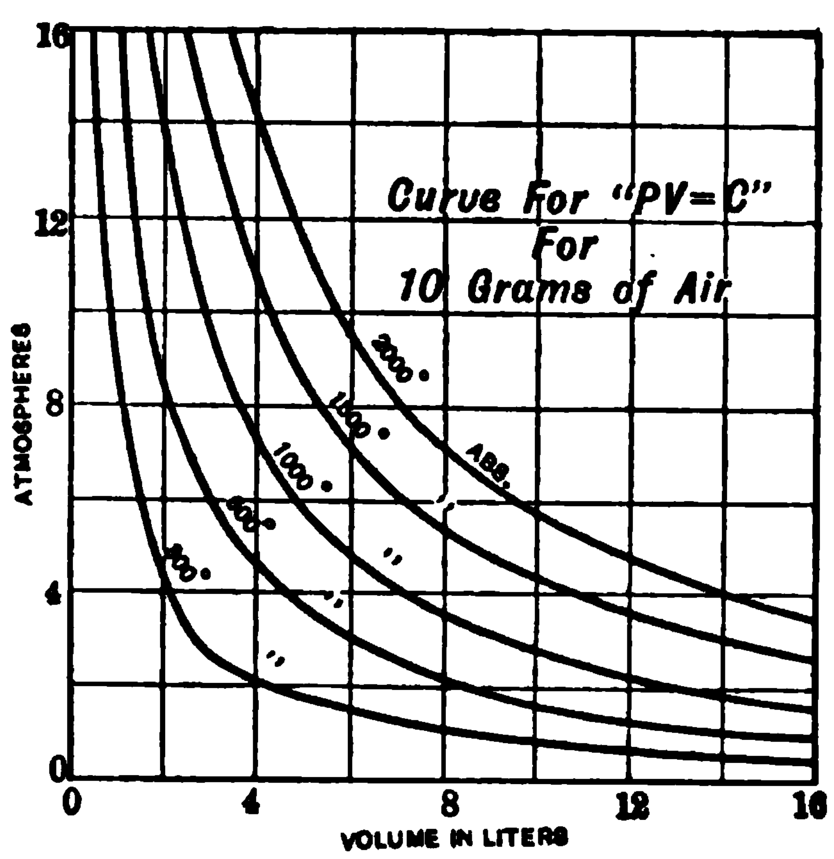


Fig. 187.—Isothermal curves for air.

distinguished by labeling it with the temperature belonging to it, as shown for air in Fig. 187. Any curve showing the relation between the *pressure* and *volume* of a substance under the condition T = const. is called an *isothermal* curve. We accordingly conclude that isothermal curves for a perfect gas are rectangular hyperbolæ, and that isothermal curves for real gases approximate to rectangular hyperbolæ, the approximation being closer at high temperatures.

283. Molecular Energy and Temperature.—As we have seen (§221) ordinary gases very approximately obey Boyle's law, $PV =$ constant for constant temperatures, and PV increases as the temperature increases. Also, according to the kinetic theory of gases (§227), for a simple ideal gas $PV = \frac{Mv^2}{3}$ which will be constant if the average random undirected kinetic energy per molecule is constant, and will increase in proportion to the average molecular energy $\frac{1}{2}Mv^2$. From these two statements for a real and an ideal gas it is natural to conclude that the temperature of a real gas is, at least approximately, proportional to the kinetic energy of molecular motion, and even to extend this analogy to liquids and solids where it has not the same justification. While the proportionality of mean molecular kinetic energy to temperature turns out to be very closely true for gases, and is a very useful and instructive hypothesis, nevertheless the complicated structure of real molecules (as compared with those of the ideal gas) shows us that the hypothesis must not be taken too literally.

CALORIMETRY

284. Unit of Heat.—Calorimetry is the process of measuring quantities of heat. Obviously the first thing to be decided upon is the unit in terms of which to measure, and though, as has been said, energy units may be used, it is often more convenient to use a unit defined in terms of *heat phenomena* only.

In looking for a purely thermal unit of heat it is natural to pick out some effect which heat produces, and agree that the heat unit shall be such an amount of heat as will produce a *specified* amount of this effect in unit mass of a *standard substance*. The *specified effect* agreed upon is a change

in temperature of 1°C., and the *standard substance* is water. To be exact the particular degree must be specified; hence we shall define the unit of heat as *that quantity of heat which will raise the temperature of* 1 *gram of water from* 14½ *to* 15½°*C.* This is called the Calorie, or Cal_{15}.

The relation of this thermal unit to the unit of mechanical energy has been found by experiments which will be described later (§340). These show that if the "mechanical equivalent of heat," that is, the number of work units equivalent to one heat unit, be denoted by the letter J,

$$J = 4.187 \times 10^7 \frac{\text{ergs}}{\text{calorie}}$$

Sometimes a *mean calorie* is also specified. This is one one-hundredth of the heat required to change 1 gram of water from 0° to 100°C. It is about equal to 1 cal. Sometimes the "large calorie" equal to 1000 calories, is used as a unit, and in engineering practice (in English-speaking countries) the "British thermal unit" (B. T. U.) is employed and is equal to the heat required to raise the temperature of 1 lb. of water 1° Fahrenheit. From the relation of the pound to the gram and the Fahrenheit to the Centigrade degree, it follows that:

1 B. T. U. = 252 Cal.

In British Thermal units and foot pounds J is 778 ft.-lbs./B.T.U.

The most common method of measuring quantities of heat in calories is by the "method of mixtures," which consists in transferring the quantity of heat to be measured to a known mass of water and observing the resulting rise of temperature of the water. The heat may be transferred to the water in many ways —for example, by dropping a piece of hot copper into the water, by pouring some hot liquid into it, or by passing steam into it. It is of course simplest to use the water at about the temperature for which the calorie is defined, as in that case the number of grams of water used multiplied by the number of degrees rise in temperature will give at once, to a first approximation, the number of calories which have been added.

285. Specific Heat.—If two different masses of water are exposed for the same length of time in just the same way to a steady source of heat, it will be found that the temperatures of the two will have risen inversely in proportion to their masses. If the same masses of copper be treated in the same way it will be found that the rise in temperature will be more than ten times as great,

but again inversely proportional to the masses. From this we conclude that the temperature effect of a given heat agent acting on a body for a given time depends on the mass of the body and on a factor which differs for different substances, and which is called the *specific heat*. The *specific heat of a substance is defined as the number of calories required to raise the temperature of* 1 *gram of the substance* 1°*C*. The symbol for specific heat is s. To be exact, the particular degree must be specified because the specific heat varies with the temperature, for example the number of calories required to raise 1 gram of a substance from 0° to 1° is different from the number required to raise it from 49° to 50°. For most purposes, however, and for not too large temperature differences, say from 0 to 100°, it is not necessary to consider the *variation* in specific heat, and it is customary to speak of *the* specific heat, meaning the mean value within the range considered.

The *heat capacity*, S, of a *body* (of any mass and variety of parts) is the number of calories required to raise its temperature 1°C. at the mean temperature t. This will evidently depend on the masses and specific heats of the various parts of the body, and if m_1, m_2, m_3 and s_1, s_2, s_3, stand for the masses and corresponding specific heats of the parts, we have

$$S = m_1s_1 + m_2s_2 + m_3s_3 + \cdots \text{ etc.}$$

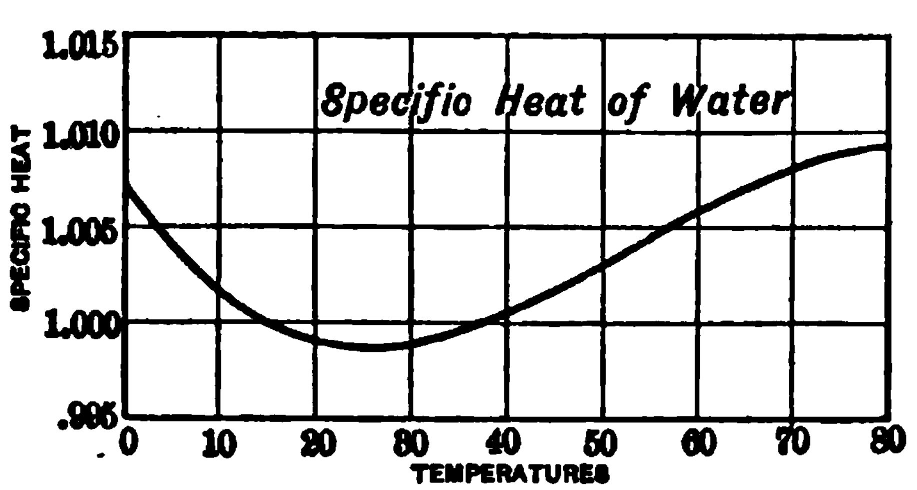


Fig. 188.—Variation of specific heat of water with temperature.

286. The Variation of the Specific Heat of Water.—The common occurrence of water and its physical and chemical characteristics make it extremely useful in heat measurements, hence a knowledge of its specific heat at various temperatures is of importance. The specific heat of water is, of course, unity at the temperature for which the calorie is defined (§284). At other temperatures it may be either greater or less than unity. The first satisfactory study of the variation of the specific heat was that of Rowland in 1878; combined with later

work, it shows that the specific heat diminishes with rising temperature, reaching a minimum between 25° and 30°C., as shown in Fig. 187. The mean value of the specific heat from 0° to 100°C. differs very little from 1.

237. Method of Mixtures.—Returning now to a more detailed consideration of the method of mixtures, there are in practice several additional points to be considered, as can best be seen by discussing a particular form of apparatus shown in Fig. 189. In the first place the water must be held in some vessel C, containing a stirrer and a thermometer, and called a *calorimeter*. Into this some of the heat will pass, raising its temperature. Moreover, some heat will pass out of the water and containing vessel during the operation and will, therefore, fail to produce its proportionate temperature change. To take account of the first effect we must know the heat capacity of the calorimeter. The second effect, loss of heat to the surroundings, necessitates what is called the *cooling* or *radiation correction*. Neglecting this correction for the moment we can write the fundamental equation of the mixture calorimeter thus

Fig. 189.—Calorimeter for method of mixtures.

$$H = (s'm + \Sigma s_1 m_1)(t_2 - t_1),$$

which expresses the fact that the heat added (H) equals the heat gained by the calorimeter and water. In this expression s', s_1 and s_2, etc., are respectively, the mean specific heats of water and of the various materials of the calorimeter, m the mass of the water, $\Sigma s_1 m_1$ the heat capacity of the calorimeter and stirrer, etc., t_2 the final and t_1 the initial temperature of the calorimeter and water.

Unless special precautions are taken the loss of heat to the surroundings, which is largely due to convection (§323) rather than radiation, is relatively great, and the cooling correction is a very important one. In general it can be reduced by protecting from air currents, polishing the ex-

posed surface of the calorimeter, surrounding it with a constant temperature enclosure, and arranging matters so that t_1 and t_2 are respectively slightly below and above the temperature of the enclosure. In some cases it is impossible or inconvenient to use water as the calorimetric substance, in which case some other liquid or solid of known specific heat may be used.

288. Application of Method of Mixtures.—The method of mixtures may be used for different purposes according to the source of the heat H which is to be measured. One important use is in determining the specific heat of substances. For this purpose a known mass M of the substance is heated to a tempera-

TABLE 5
SPECIFIC HEATS
(Calories per Degree C. per Gram)

Substance	Specific heats	Temperature C.
Alcohol (ethyl)	0.548	0°
Aluminum	0.219	15 to 185
Aluminum	0.0093	−240
Brass	0.090	0
Copper	0.0936	20 to 100
Copper	0.00036	−250
Diamond	0.113	11
Diamond	0.0003	−220
Glass (flint)	0.117	10 to 50
Gold	0.0316	0 to 100
Granite	0.19 to 0.20	0 to 100
Graphite	0.160	11
Graphite (acheson)	0.0573	−79 to −190
Ice	0.502	−21 to −1
Iron	0.119	20 to 100
Lead	0.0305	20 to 100
Lead	0.0143	−250
Mercury	0.0333	20
Mercury (solid)	0.00329	−40 to −75
Nickel	0.109	18 to 100
Platinum	0.0323	0 to 100
Quartz	0.174	0
Silver	0.0559	0 to 100
Sodium	0.2433	−83 to −190
Tin	0.0552	19 to 99
Turpentine	0.420	18
Sea water	0.980	17
Zinc	0.0935	0 to 100
Zinc	0.0017	−240

ture t (above or below t_1) and added to the calorimeter and water. The temperature of the calorimeter and of the mass M will then equalize and, if we call t_2 the final temperature of the mixture, the heat H added to the calorimeter is the heat lost by the mass M in changing from the temperature t to t_2, which, from the definition of specific heat, is equal to $sM(t-t_2)$ if s is the

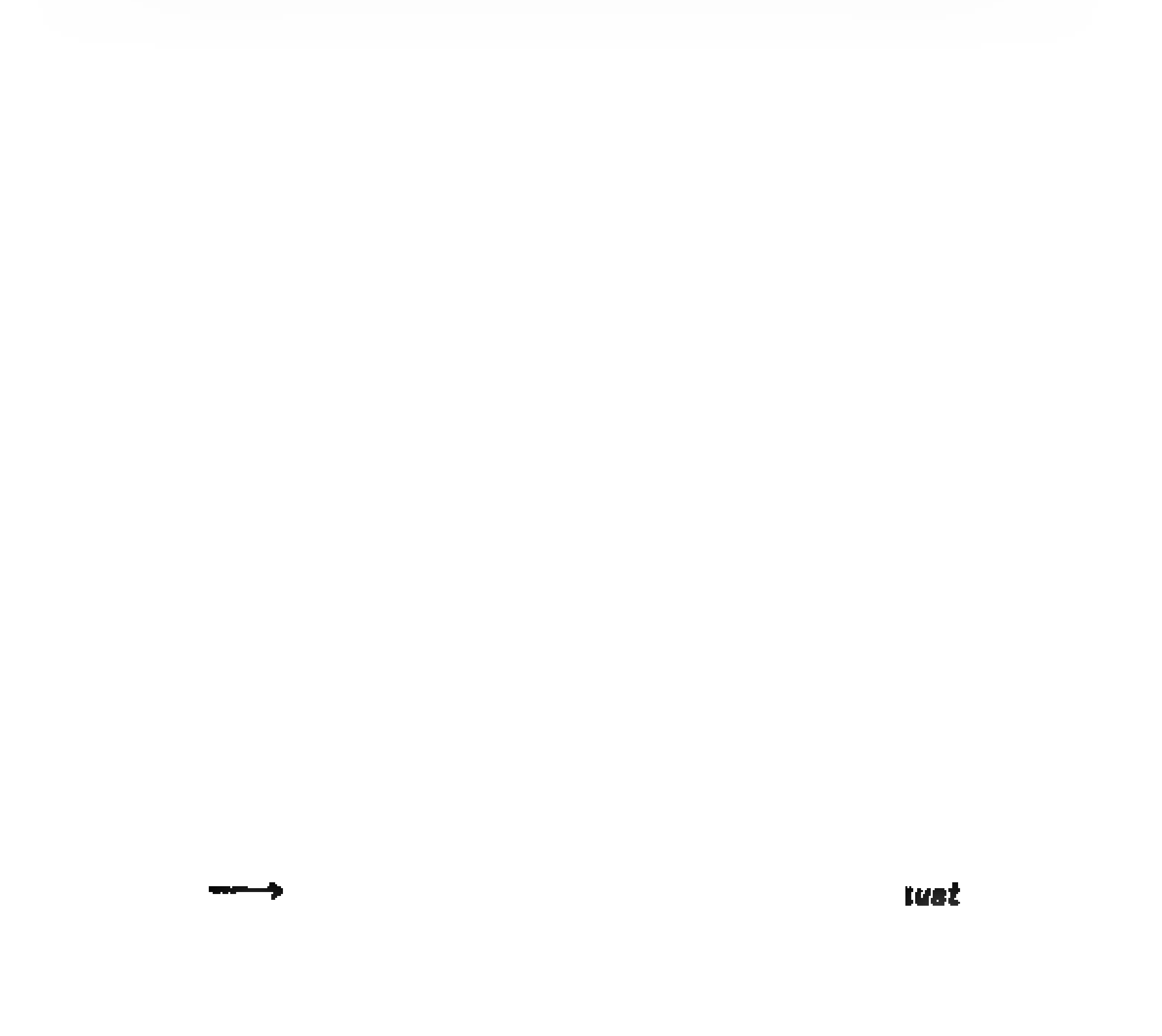


FIG. 190.—Continuous flow calorimeter for measuring heat of combustion of gas.

mean specific heat of the substance M in the interval t to t_2. We then have:

$$H = Ms(t-t_2) = (ms' + \Sigma m_1 s_1)(t_2 - t_1)$$

from which s may easily be computed.

Sometimes it is advisable to keep the hot body from direct contact with the water by putting it in an inner vessel having thin walls of good conducting material.

The method of mixture is also used to determine heats of fusion (§305) and evaporation (§312) as well as the amount of heat

developed or absorbed in various chemical reactions. In such cases the operation consists in fusing, or condensing, or combining, as the case may be, known masses of material inside the calorimeter (in the inner vessel above referred to), and special forms of calorimeters, called combustion calorimeters, bomb calorimeters, etc., have been developed for these purposes.

289. Method of Continuous Flow.—A second method for measuring quantities of heat is the method of "continuous flow," illustrated in Fig. 190, in which a steady stream of the calorimetric substance (usually water from a reservoir) at a constant temperature is allowed to flow past the point at which heat is being set free, in such a manner that all of the heat is absorbed by the stream of water. The temperature of the stream of water is, of course, higher after the heat has been absorbed than before, and if the rate of liberation of heat is constant this temperature difference will be constant and (neglecting external losses as before) the number of calories liberated in a time T (*since it does nothing but heat the water*) will be equal to the number of grams of water W which has flowed past in time T, multiplied by the number of degrees rise in temperature $(t_2 - t_1)$, and by the specific heat of water s',

or,
$$H = Ws'(t_2 - t_1)$$

This method is especially useful in determining the heats of combustion of gas and liquid fuels by means of which a steady rate of combustion and hence a steady liberation of heat can be maintained. This method can also in a sense be *reversed* by generating the heat mechanically or electrically, that is, by converting measured amounts of mechanical or electrical energy completely into heat, which is absorbed by a stream of fluid whose specific heat is to be determined. The above equation then becomes:

$$H = Ms(t_2 - t_1)$$

where H is known (from mechanical or electrical measurements) in energy units, M is the mass of fluid flowing past in time T and s is its specific heat which is determined by this equation in mechanical units. In this form the method has been used by Barnes to measure the specific heat of water and mercury, and it is capable of giving very accurate results.

290. A third method of measuring quantity of heat is the "method of latent heats," in one form of which the heat to be measured is used to melt a measurable amount of ice. This necessitates a knowledge of the amount of heat required to melt 1 gram of ice (heat of fusion of ice, §305), but has the advantage that the calorimeter remains at a fixed temperature, 0° C. The most common instrument of this type is the Bunsen ice calorimeter. A second form is the Joly steam calorimeter, in which the order of temperatures is reversed, and the amount of heat required to raise M grams of a substance from a temperature t to the temperature of steam, say 100°C., is determined from the weighed amount of steam which is condensed to

supply this heat. A knowledge of the heat liberated in condensing 1 gram of steam is, of course, necessary. This is a very convenient and reliable method.

291. The Specific Heat of Gases.—In the previous discussion of specific heat we have neglected one factor which, as we have seen in §279, becomes very important as soon as we consider gases, namely, the expansion which usually accompanies rise in temperature. If a gas is confined in a cylinder with a movable piston, as, for instance, in a bulb with a mercury plug in an attached capillary tube, Fig. 172, and is heated, it will, as we have already noted, expand and push out the mercury plug. The outside of the plug is acted upon by the pressure of the air which opposes its motion outward by a force equal to the product of the pressure and the cross-section of the tube. Overcoming this force through a given distance means doing work, called the *external work of expansion*, and this work has evidently been done by the expanding gas.

Looking at the matter from the standpoint of the kinetic theory, we should say that, before the confined gas was heated, the impact of gas molecules on the inner end of the mercury plug (at rest) was balanced by the impact of air molecules on the outer end, but that an increase in temperature of the gas meant more and harder impacts on the inner end, thus destroying the equilibrium and causing the plug to move. The *moving* plug would, on the average, hit the outside molecules harder than it had previously done when at rest; hence it would increase the velocity of these molecules and add kinetic energy to them. The work done by the expansion consists in a transfer of kinetic energy from the gas molecules inside to air molecules outside.

Thus, if heat energy imparted to the confined gas causes expansion some energy will be, by this expansion, taken out of the gas, and this is a possible disposition of part of the added energy quite separate from those considered in §262. Hence we can see that to raise the temperature of a gas with the volume kept constant must take an amount of energy different from that required if expansion against pressure is allowed, not only because in the second case the internal potential energy may be increased but because external work is done. In other words, the heat added to a gas (or any body) is equal to the increase in *internal* kinetic and potential energy plus the *external* work done.

The amount both of the internal and of the external work will evidently depend on the amount of expansion. Since the

increase in volume of gases per degree rise in temperature is very much greater than that of solids or liquids, the external work is greater. If the volume is not kept constant it may be allowed to vary in many mays, the most important being such an increase in volume that the *pressure remains constant*. Hence we have the specific heat of a gas at constant volume, s_V, and the specific heat at constant pressure, s_P, defined as the heat necessary to raise the temperature of 1 gram of the gas 1°C. under the condition of constant volume or constant pressure respectively. From what has been said it is evident that s_P must be greater, in general considerably greater, than s_V.

TABLE 6

SPECIFIC HEATS OF GASES AND VAPORS

Substance	Temperature	Specific heats		s_P/s_V
		s_P	s_V	
Alcohol (ethyl)	108–220	.453	.400	1.133
Air	0–100	.241	....	1.402
Argon	20– 90	.123	....	1.667
Benzine	34–115	.299	.214	1.397
Carbon dioxide	15–100	.2025	....	1.299
Chlorine	16–343	.113	....	1.336
Chloroform	27–118	.144	.125	1.152
Ethyl ether $(C_2H_5)_2O$	25–111	.428	....	1.024
Helium	....	....	....	1.63
Hydrogen	12–198	3.409	....	1.408
Mercury vapor	310	....	....	1.66
Nitrogen	0	.235	....	1.41
Oxygen	20–440	.242	....	1.398
Water vapor	100	.442	....	1.33

The measurement of s has been most accurately made by means of the steam calorimeter (§290), a known mass of the gas being enclosed in a metallic bulb, and the weight of steam condensed in raising it from t°C. to 100°C. being determined; a correction must then be made for the thermal capacity of the bulb. s_P is usually measured by passing a stream of heated gas

through a calorimeter. According to Regnault and later observers s_P for most gases varies only slightly with pressure, while s_P for air is almost independent of the temperature, but for CO increases very markedly with temperature. s_V for air and CO_2 increases with the density of the gas. The value of s_V has not been determined directly for many gases, but the value s_P/s_V can be readily deduced from the velocity of sound in the gas (§ 587).

292. The Free Expansion of a Gas.—We have already seen that, if there are forces between molecules and atoms, when a gas expands there will be a change in the potential energy of its molecules (and perhaps of its atoms), since the average distance between molecules will increase. Work done against internal forces in this way is called the *internal work* of expansion to distinguish it from the *external work* done against the pressure confining the gas.

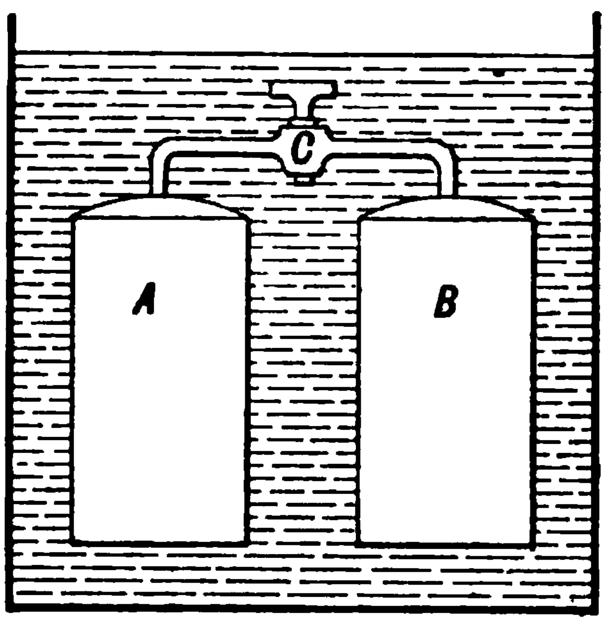


FIG. 191.—Illustrating Joule's study of the "free expansion" of gases.

Gay-Lussac and later Joule attempted to measure the internal work by the method of free expansion, Fig. 191, in which gas was confined at some considerable pressure in the vessel *A* and allowed to expand quickly through a cock *C* into B which had been highly exhausted. *A*, *B*, and *C* were in a vessel of water whose temperature was measured. Since expansion occurred into a vacuum it was "free" (no opposing pressure) and hence, on the whole, no external work was done; but if there were any internal work done, it must have resulted in a change in temperature of the gas. For if the internal potential energy increases, the kinetic energy must decrease by an equal amount, that is, the temperature of the gas must fall, and *vice versa*. Joule did not measure the temperature of the gas itself, but that of the water, whose large heat capacity so masked the effect that his results merely indicated that the *internal work of expansion is small.*

293. Temperature of Gas in Motion.—If a gas at high pressure and ordinary temperature is allowed to escape into the atmosphere through a fine tube (Fig. 192) in which it acquires a high

velocity of flow, very marked cooling effects will be observed where the velocity of flow is greatest, though the total energy of the moving gas is practically the same as that of the gas at rest. The explanation is that part of the energy of the random undirected motion of the particles which determines the temperature has become temporarily energy of directed motion in the stream. But if the gas is caught in a large receiver and allowed to come to rest its temperature will be found to be slightly higher than before expansion. If A, B, and C Fig. 191 are placed in separate vessels it will be found that the expansion lowers the temperature of A, and raises that of B an equal amount. The reason for this is that the gas moving out of A corresponds, with respect to the gas remaining in A, to the piston moving *away from* the gas in Fig. 172, hence the gas remaining does work on the gas which is going, and the one loses and the other gains heat of equal amount.

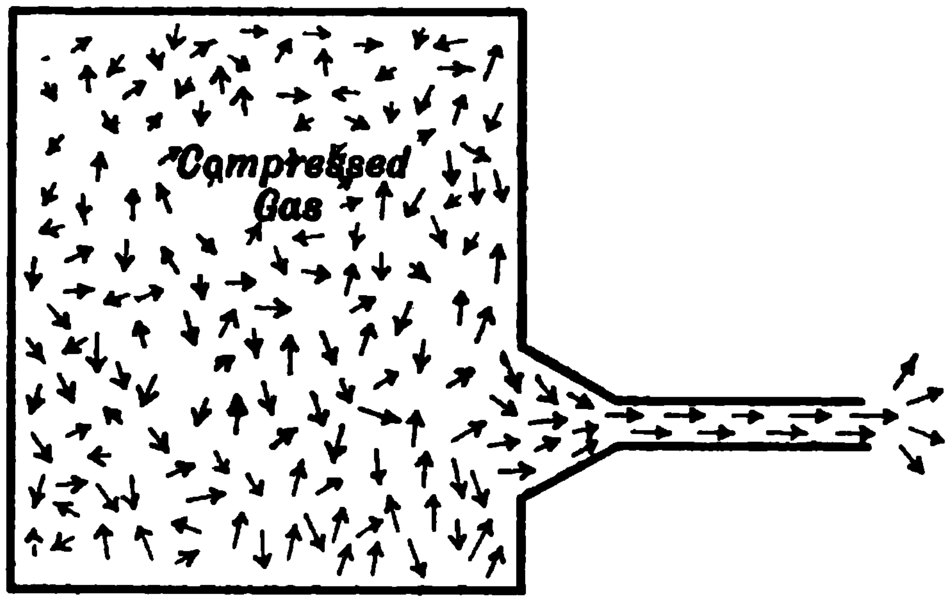


FIG. 192.—The change from undirected to directed molecular motion in an escaping gas.

294. The Difference Between the Two Specific Heats.—From the definition of s_P and s_V in §291 and from the statements made in §292 we see that if we denote the external work of expansion by W_e and the internal work of expansion by W_i

$$s_P = s_V + W_e + W_i$$

The expansion is that necessary to maintain P constant while t rises 1°C., and the two "works" must be expressed in heat units. For gases which are not approaching liquefaction, that is, for example, for O, H, N, and air, at ordinary temperatures, the internal work of expansion is so small that it may usually be neglected, so that if P is the constant pressure, and ΔV the

change in volume per degree per unit mass, and J the mechanical equivalent of heat, then the external work $=\frac{P\Delta V}{J}$

and $$s_P - s_V = \frac{P\Delta V}{J}$$

Also, $$PV = RT \qquad \text{approximately,}$$

and $$P(V+\Delta V) = R(T+1)$$

Hence $$P\Delta V = R$$

and $$s_P - s_V = \frac{R}{J}$$

This equation was used by Robert Mayer in 1842 to make the first computation for J, the other quantities being determined by experiment.

295. The Ratio of the Two Specific Heats.—From what has just been said, the ratio of the two specific heats is evidently:

$$\frac{s_P}{s_V} = \frac{s_V + \frac{W_i}{J} + \frac{R}{J}}{s_V}$$

Also, according to the kinetic theory, $s_V =$ increase in molecular kinetic energy + increase in atomic energy, per degree; of which the first, which we shall denote by E_m, represents the increase in temperature, the second the increase in energy *inside* the molecule which we shall represent by E_a. From §227

$$PV = \frac{Mv^2}{3} = \frac{2}{3} \cdot \frac{1}{2} Mv^2 = RT.$$

Hence $$E_m = \frac{1}{T} \cdot \frac{1}{2} Mv^2 = \frac{3}{2} R$$

so that $$s_V = \left(\frac{3}{2}R + E_a\right)\frac{1}{J}$$

and $$\frac{s_P}{s_V} = \frac{\frac{3}{2}R + E_a + W_i + R}{\frac{3}{2}R + E_a}$$

If E_a and W_i are both relatively small, as we should expect them to be simple monatomic gases which apppoximately obey Boyle's law, for example, argon, then

$$\frac{s_P}{s_V} = \frac{5}{3}, \text{ approximately.}$$

On the other hand if all the other terms are negligible compared with E_e. as we might expect for gases with very complicated molecules, for example ether,

$$\frac{s_P}{s_V} = 1 \text{ approximately.}$$

Both results are in agreement with the values given in § 291.

296. Expansion Against Pressure.—Let a gas be forced through a small aperture in such a way that the pressures before and after passing the opening are maintained constant. A possible way of doing this is shown in Fig. 193, in which the pistons both move to the right as the gas passes through, and the external forces upon them are constant.

Let P_1 = pressure of gas before expansion.
V_1 = specific volume before expansion.
P_2 = pressure of gas after expansion.
V_2 = specific volume after expansion.

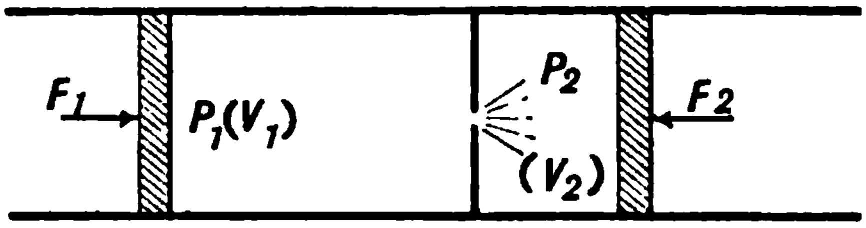


FIG. 193.—The Joule-Kelvin porous plug experiment. Unbalanced but not "free" expansion.

Then the external work done upon the gas by the first piston while unit mass is passing is P_1V_1 (§195), the external work done by the gas after expansion upon the second piston is P_2V_2, and $(P_1V_1 - P_2V_2)$ is the net amount of external work, W_e, done *on* the gas, and this may be either positive or negative. The apparatus is supposed to be so made that no heat can enter or leave the gas during the operation and the temperature of the gas is observed before and after expansion.

Let W_i again represent the internal work done against the forces between molecules, and Δt the observed change in temperature of the gas. Then the sum of the external and internal work must be equal to the change in the kinetic energy of the molecules plus any change in the energy inside the molecules, and this, as was just shown in §295, is s_V. We therefore have the equation

$$\Delta t s_V = W_e + W_i$$

If the gas strictly obeyed Boyle's law and if there were no temperature change, P_1V_1 would be equal to P_2V_2, that is W_e would be zero and hence, by the above equation, the internal work would be zero. As a matter of fact, with O, N, and CO_2 a cooling is observed, with H at ordinary temperatures a heating, and from these observations, combined with the value of the specific heat s_V and the variation of PV, the internal work, which Joule could not detect, can be computed. The results indicate in all cases *molecular forces of attraction.* To avoid false cooling effects due to mass motion of the gas (§293), it is customary, following the plan of Lord Kelvin, to use many fine openings—that is a *porous plug*—hence the experiment is known as the "porous plug experiment."

297. Relations between Specific Heats.—In view of the complicated nature of molecular structure as already outlined, it is evident that no simple relations are to be expected between the specific heats of different substances or of the same substance at various temperatures, and the following general statements must suffice.

The best known attempt to express a relation between the specific heats of substances is the so-called "law" of Dulong and Petit (1819), which states that the "product of the specific heat by the atomic weight is the same for all solid elements." But this statement and its extension to compound substances are only rough approximations.

In general the specific heat of substances in the liquid state is much greater than in the solid (two times as great for water, ten times for mercury), while s_P for the gaseous state is about the same as that of the solid. The change of specific heat from solid to liquid is in most cases smaller with metals than with non-metals.

With most substances, solid, liquid or gaseous, the specific heat increases with rise of temperature, though the change is small for solids with the exception of carbon, boron and silicon. This variation of the specific heat may, according to §262, be due to several causes, such as an increase in the relative amount of kinetic energy inside the molecule and atom, increase in the number of free electrons, or an increase in the potential energy of molecular groups or groups of atoms. The change of molecular grouping is probably the chief cause of variation. The fact of variation with temperature shows at once that Dulong and Petit's "law" cannot be a general one.

Quite recently Nernst has extended the measurement of specific heats down to 23° abs. (−250°C.), and has shown that the specific heat of all the substances examined decreases very greatly at extreme low temperatures (Table 5). He has also developed new relations between the specific heats of elements and compounds which promise to be of great importance.

298. Heats of Combustion.—A very important use of calorimeters is in measuring heats of combustion of fuels, that is, the heat liberated by the burn-

ing (in air or oxygen) of 1 gram of coal, wood, oil, gas, etc. Such fuels are the source of the larger part of the available energy of the world, and a knowledge of the energy available per unit mass of the fuel is, of course, of great importance to the engineer. The method of mixtures is usually used for solid fuels, especially with one form of apparatus called a "bomb calorimeter," in which a weighed amount of fuel is enclosed with compressed oxygen in a steel bomb and ignited electrically, the bomb being in the water of the calorimeter. For liquid fuels and gases a method of continuous flow (§289) is also very much used. The heat of combustion is usually expressed in calories per gram, or B. T. U. per pound of fuel.

TABLE 7

HEATS OF COMBUSTION. (CALORIES PER GRAM.)

Substance.		Substance.	
Alcohol (ethyl)	7183	Gas (coal gas)	5800–11000
Alcohol (methyl)	5307	Gas (illuminating gas)	5200–5500
Benzine	9977	Gunpowder	730
Carbon (diamond)	7860	Hydrogen	34100
Carbon (graphite)	7800	Petroleum (Am. crude)	11100
Coal (anthracite)	7600–8400	Wood (beech)	4168
Coal (bituminous)	6100–7800	Wood (oak)	3990
Coal (coke)	7000	Wood (pine)	4420

CHANGE OF STATE

299. Change of State.—The most marked changes in the physical properties of bodies occur when they change from the solid to the liquid or gaseous state.

The change from the solid to the liquid state is called *fusion* or *melting*, the reverse change, *freezing*.

The change from the liquid to the gaseous state is called *vaporization*, the reverse change *condensation*.

The change from the solid directly to the vapor state is called *sublimation*, the reverse change, *condensation*.

Each of these changes involves a rearrangement of the molecules with respect to each other, and perhaps a rearrangement of the atoms and electrons forming the molecules. Vaporization and sublimation also involve a very great increase in the average distance separating molecules. Rearranging and separating the molecules will involve an increase in potential energy in passing

from the solid to the liquid and to the gaseous state, while any change in volume will involve doing external work (§292); hence energy must be added to the body to bring about the change. Conversely, when a vapor condenses or a liquid solidifies, a certain amount of energy must be taken away from it. As groups of liquid molecules "settle down" into the solid arrangement, some of their potential energy becomes kinetic and is given up to the surface on which freezing occurs.

300. Fusion.—If a crystalline solid is heated while acted upon by a constant pressure, it will begin to melt at a definite temperature called the normal fusing-, or melting-point, and the entire mass will remain at this temperature until it is all melted. To determine this temperature, a thermometer bulb of some kind protected by a metal or porcelain tube, may be put in the mixture of solid and liquid, as in Fig. 194. Or, a thermometer may be placed in a molten substance which is allowed to slowly lose heat; the temperature will fall until solidification begins after which it will remain constant, while potential energy (heat of fusion) is being given up, until solidification is complete. The constant temperature is the *freezing-point* of the substance

The freezing-point of water is found by immersing a thermometer in a mixture of pure ice and water, carefully protected from gain of heat from the outside. As has been stated, the *freezing-point of water under one atmosphere pressure* is one of the fixed points of thermometry.

Fig. 194.—Method of determining a freezing-point

301. Effect of Pressure on Fusion.—The normal melting-point of pure substances depends upon the pressure under which fusion occurs—*increase* of pressure *raising* the melting-points of those substances which expand on melting and *lowering* the melting-points of those substances which contract on melting. This relation between change of melting-point with pressure, and change of volume on fusion, was deduced first from theoretical considerations by James Thomson. The curve obtained by plotting melting-points and the corresponding pressures on the *Pt* diagram is called the fusion-curve (Fig. 200), and represents

the conditions under which the solid and liquid can exist in equilibrium in contact. The change in the melting-point with pressure is in general not large, being for water 0.0072°C. *decrease* in temperature for each atmosphere increase in pressure. As a result of the work of Bridgman and Tammann it is now known that water can exist in five different solid forms, which can be in equilibrium with ordinary ice or liquid water under conditions roughly given by the dotted curves from the end of I, Fig. 200. There is a minimum freezing point for water at −22°C. under a pressure of 2500 atmospheres, while the melting-point of one form of ice has been followed up to 80°C. at a pressure of 20,000 atmospheres.

TABLE 8

TABLE OF MELTING-POINTS

Substance.	Melting-point.	Substance.	Melting-point.
Aluminum	657°C.	Nickel	1452
Copper	1083	Platinum	1753
Gold	1063	Silver	960
Iridium	2290	Tin	232
Iron	1505	Tungsten	3270
Lead	327	Zinc	418
Mercury	−38.8		

The lowering of the freezing-point of water with pressure may be strikingly illustrated and has important consequences. Two blocks of ice at about 0°C. will freeze together if two faces are pressed together; snow compressed in a cylinder becomes a clear transparent mass of ice, and snow at about 0°C. can be pressed by the hands into a hard snow ball. If a wire supporting a weight is looped around a block of ice, it will slowly melt its way through the ice, which freezes again above it. A further illustration is the fact, well known to skaters, that ice is more slippery when near 0° than when many degrees below zero. In all these cases the pressure applied, which may be very considerable at certain points, lowers the melting-point, and, if the initial temperature of the snow or ice is not too low, some of it will melt, only to freeze again when the pressure is relieved. Thus there would be a film of water between the skate runner and the ice. If the ice or snow is too cold, the pressure will not lower the melting-point below this initial temperature and no melting will occur.

The same ideas apply to the "packing" of snow on roads, and on a larger scale to the formation of glaciers by the compression, melting, and *regelation* of snow in mountain valleys. The subsequent flow of glaciers down the valleys is due in part to the effect here discussed, the ice melting at the points of greatest pressure, the water immediately flowing down hill a little, thus relieving the pressure, and then freezing again.

302. Crystalline and Amorphous Solids.—The sharp change from solid to liquid at a definite temperature, which we have been discussing, is characteristic of solids which have a definite crystalline structure. Solids which have not such a structure,

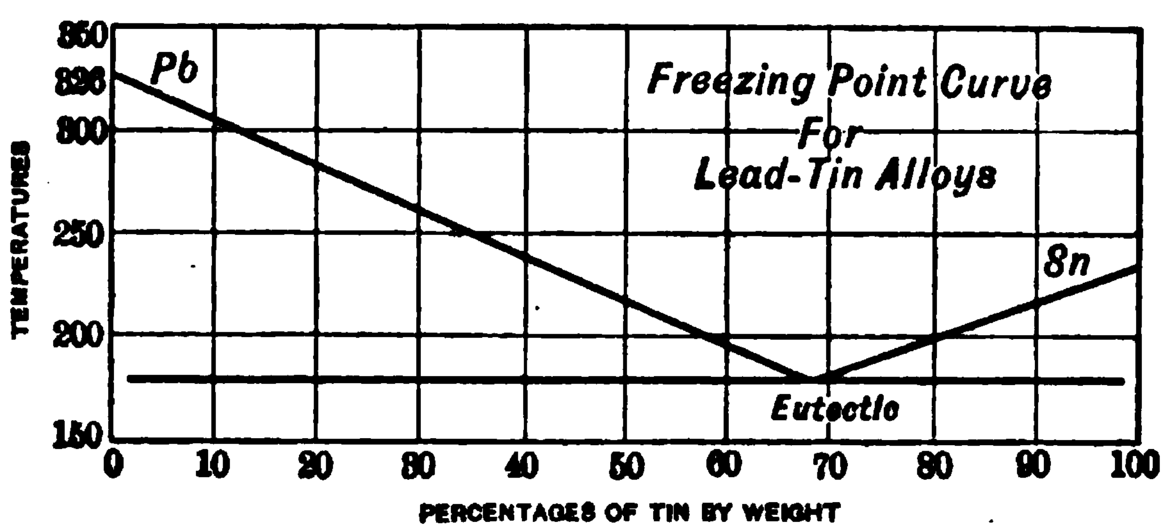


FIG. 195.—Freezing-points of alloys. Upper line, beginning of freezing; lower line end of freezing; eutectic has a sharp freezing-point.

called amorphous solids, of which fats, waxes, glass and most alloys are examples, change gradually from one state to another, that is, gradually soften throughout the entire mass, while the temperature rises slightly, there being no definite "melting-point." Amorphous solids are in general mixtures. Some alloys, however, are definite *compounds*, having marked crystalline structure and very definite melting-points. The freezing-point curve for a simple group of alloys is shown in Fig. 195.

303. Change of Volume on Freezing.—Most substances contract on freezing, the solid sinking in the liquid. The fact that iron, bismuth and antimony and some alloys such as type-metal (lead, antimony, and tin) expand on solidifying is valuable industrially, since, when cast, they take a particularly sharp impression of the mold. The expansion of water upon freezing is responsible for the bursting of water pipes, the bursting (and hence death) of plant cells, and the splitting of trees and rocks. Very carefully dried seeds may be put in liquid air without injury,

but the presence of the slightest trace of moisture will result in killing the seeds.

304. Freezing Point of Solutions.—The fact that a dilute solution, such as sea water, has a lower freezing-point than the pure solvent, and that the lowering of the freezing-point of dilute solutions is approximately proportional to the amount of substance dissolved has been known for a long time. The depression of the freezing-point per gram-molecule of salt dissolved in 100 grams of solvent, calculated from observations on dilute solutions, is called the *molecular lowering of the freezing-point.* Later work shows that the freezing-point of a given solvent is lowered the same amount by many different salts when dissolved in proportion to their molecular weights, while other salts will produce a depression two or three times as great.

According to the dissociation hypothesis, abnormally large depressions are due to the breaking up, or dissociation, of molecules into parts, while abnormally small depressions are due to the grouping together of molecules. Thus common salt in water apparently dissociates into Na and Cl giving a solution which conducts electricity readily, and producing a molecular lowering of the freezing point of about 3.6°. If the temperature of a given dilute solution is lowered beyond the freezing point corresponding to its saturation, the pure solvent only will begin to freeze out of the solution, which becomes, therefore, more concentrated, until, on continued cooling, a certain definite concentration is reached (depending upon the pressure) when the entire mass freezes as a mixture of the two solids. This mixture is called a *cryohydrate*. The corresponding mixture in case of alloys, having a minimum melting-point as compared with other percentage compositions, is called an *eutectic*. (Fig. 195.)

305. Heat of Fusion.—The *heat of fusion* of any substance is defined as *the number of calories required to convert one gram of the solid at the melting-point into liquid at the same temperature.* Heats of fusion are usually measured by some modification of the method of mixtures.

Thus if M = no. of grams of melted substance used,

t_s = temperature of substance when added to calorimeter,

t_m = melting point of substance,

t_2 = final temperature of calorimeter,

t_1 = initial temperature of calorimeter,

m = mass of water used,

$\Sigma m_1 s_1$ = heat capacity of calorimeter,

s_l = specific heat of substance when melted,
s_s = specific heat of substance when solid,
L = heat of fusion,

then

$$(m+\Sigma m_1 s_1)(t_1-t_2)=M[s_s(t_m-t_2)+L+s_l(t_2-t_m)]$$

from which L may be computed.

TABLE 9

HEATS OF FUSION

(Calories per gram)

Aluminum	77
Copper	43
Ice	79.8
Lead	5
Mercury	3
Platinum	27
Sulphur	9
Zinc	28

306. Vaporization.—From the molecular standpoint, vaporization means the flying off of molecules against the forces of molecular attraction, these molecules losing kinetic energy and gaining potential energy as they leave the liquid. The more rapidly moving molecules will be the first to fly off, hence the average kinetic energy of the molecules remaining behind will be less than the initial average for the liquid, and the liquid will be *cooled* by evaporation. If the vapor is confined over the liquid, some vapor molecules will strike the surface and become liquid again, and as the number of vapor molecules per unit volume (*i.e.*, the density of the vapor) increases, the number of molecules returning to the liquid per second will likewise increase, until finally the average number returning will equal the average number leaving. Under these conditions *the vapor is in equilibrium with the liquid*. The density, and hence the pressure, of the vapor which

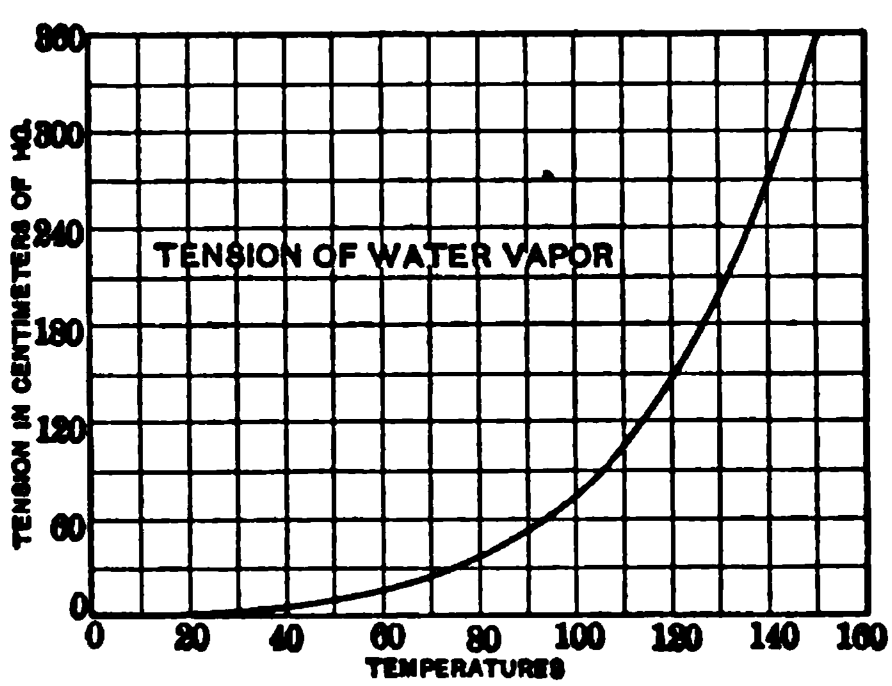


FIG. 196.—Pressure (tension) of water vapor at various temperatures.

will be in equilibrium will depend on the temperature, that is, on the average molecular velocity. A vapor in equilibrium with the liquid is said to be *saturated*, and the equilibrium pressure is called the *saturated vapor pressure* (or *vapor tension*), which for a given substance depends only upon the temperature. If the vapor is not allowed to accumulate over the liquid it will remain unsaturated, equilibrium will not be reached, and the liquid will gradually disappear by evaporation.

TABLE 10

VAPOR TENSIONS AND VAPOR DENSITIES OF WATER

Temperature.	Vapor tensions (mm. of Hg.).	Densities of saturated vapor (grams of vapor per cu. m. of saturated air).
−20° C.	0.781	0.892
−10	1.961	2.154
0	4.579	4.835
10	9.205	9.330
20	17.51	17.118
30	31.71	30.039
40	55.13	50.625
50	92.30	......
60	149.2	......
70	233.5	......
80	355.1	......
90	525.8	......
100	760	......
140	2709	......
180	7514	......
260	35760	......
360	141870	......

No general relation is known connecting the saturated vapor pressure and temperature, though many empirical relations have been found which are satisfactory in certain cases. The corresponding values of temperature and saturated vapor pressure for water are shown in Table 10 and Fig. 196. Points in this diagram indicate the physical condition of water substance; points on the curve show the conditions under which water may exist either as a vapor or liquid or both in equilibrium, as, for example, at a temperature of 140°C. and under a pressure of 270 cm. of mercury. If the pressure is increased, without suitably raising the temperature so as to reach another point on the curve,

all the vapor will be condensed, while if the temperature is increased without properly increasing the pressure, all the water will vaporize. Hence this curve, which represents equilibrium conditions, divides other conditions into two groups, an all-vapor group represented by points to the right of and below the curve, and an all-liquid group represented by points to the left of and above the curve.

307. Humidity.—The saturated vapor pressure for a given temperature is not measurably affected by the presence of gases which do not chemically combine with the vapor. When we speak of air being saturated with water vapor, what we really mean is that the vapor is saturated. The presence of air above a water surface will not influence the vapor pressure necessary for equilibrium, but will slightly influence the rate of evaporation if the equilibrium condition is not reached.

The degree of saturation of air with water vapor is of great importance in its influence upon climate, for it determines the rate at which evaporation will go on from exposed surfaces of water or from moist surfaces, such as that of the human body. Evaporation, as we have seen, causes cooling; hence the less saturated the air the greater the cooling, since evaporation will be more rapid. Thus a given summer temperature with the air dry is less oppressive than with the air nearly saturated. The effective dryness of air depends on its degree of saturation, and this is called the humidity, *absolute humidity* being defined as the mass of water vapor contained in a cubic centimeter of air at a given temperature, and *relative humidity* as the ratio of the mass of moisture actually present to the amount needed for saturation.

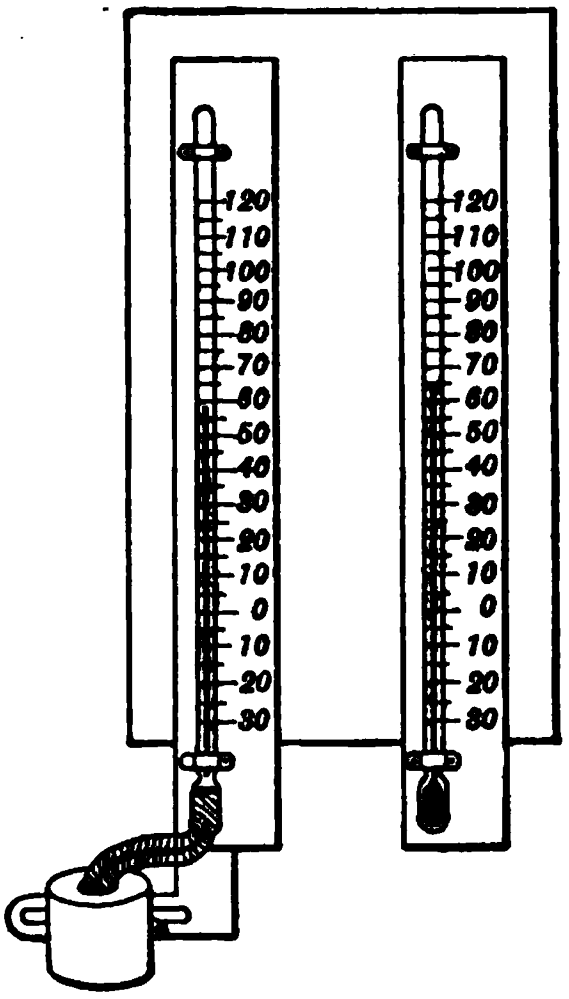


Fig. 197.—Wet and dry bulb hygrometer.

If water vapor (or air and water vapor) is heated at approximately constant pressure, without the addition of vapor, as in a hot air furnace, it expands, and therefore the mass of vapor

per unit volume decreases. At the higher temperature the density necessary for saturation is greater, however, so that for two reasons the effective dryness is increased. As might be expected, the air in houses in winter is usually far too dry for either comfort or good health.

The measurement of humidity is called *hygrometry*. The *wet and dry bulb hygrometer* (Fig. 197) consists of two exactly similar thermometers similarly exposed, except that the bulb of one is covered with a light wick kept moist by dipping in a vessel of water. Evaporation from the wet bulb will cool it as compared with the other, and the dryer the air the greater will be the difference in temperature between the two. By noting this difference and the temperature of the dry bulb, either the relative or the absolute humidity may be obtained from tables. Air must circulate freely around these thermometers in order that they should give accurate results.

Much more reliable results are obtained if the wet and dry bulb thermometers are swung rapidly through the air, this form being called the "sling psychrometer." Other methods depend upon the determination of the *dew-point*, that is, the temperature at which air would be saturated by the moisture actually in it. One form of dew-point hygrometer, Regnault's, is shown in Fig. 198. The central chamber into which the thermometer dips is cooled by the evaporation of the ether contained in it, and the temperature at which condensation first occurs on the front polished metal face is the dew-point. At ordinary room temperatures the proper relative humidity is from 60 to 70 per cent. To reach this humidity, it is necessary to evaporate water in the furnace or in the rooms themselves. If evaporation is desired in the furnace, it is evident, from what has been said above, that the *hot* air (not the *cold* air as is usually the case) should pass over the water surface in order to take up as much water vapor as possible and not be "dried" by subsequent heating. To give 70 per cent. relative humidity at 68°F. in a house 40 ft. by 30 ft. and 25 ft. high, containing about 30,000 cu. ft., requires about 22½ lbs. or 2.8 gallons of water, and usually several times this amount must be supplied per day to maintain proper conditions.

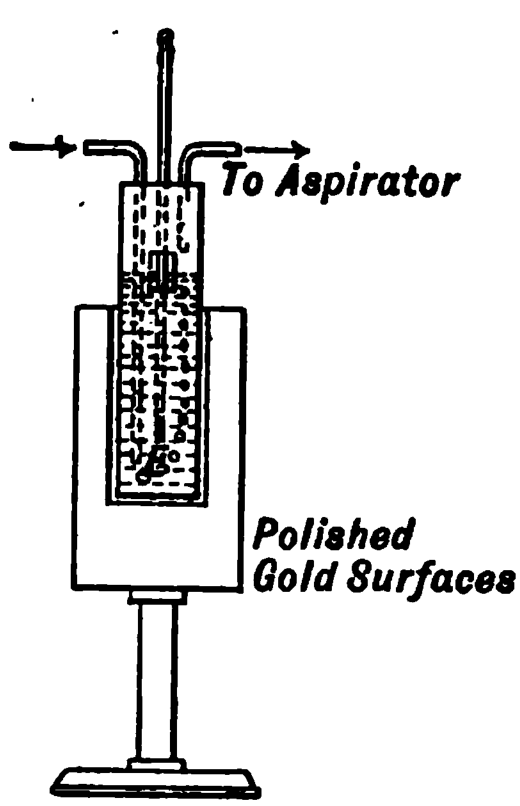


FIG. 198.—Regnault's dew-point hygrometer.

308. Boiling-point.—*The boiling-point of a liquid is the temperature at which its saturated vapor pressure is equal to the atmos-*

pheric pressure on the liquid surface. At this temperature, bubbles of vapor will form in the liquid and escape through the surface, and this formation and escape of vapor bubbles is called *boiling.* Evidently the boiling-point will be higher the greater the pressure on the surface of the liquid, and the variation of the boiling-point with pressure is the same thing as the change of saturated vapor pressure with temperature. In the determination of boiling-points the thermometer is put in the vapor rather than in the liquid, and it must be protected from gain or loss of heat by radiation and from the condensation of liquid upon it.

TABLE 11

BOILING-POINTS UNDER ONE ATMOSPHERE PRESSURE

Substance.	Boiling-points.
Alcohol (ethyl)	78.3
Benzol	80.2
Carbon dioxide	−78.2
Chloroform	61.2
Ether (ethyl)	34.6
Iron	2450
Mercury	357
Oxygen (liquid)	−182.9
Pentane	36.2

309. Effect of Pressure on Boiling-point.—The variation of the boiling-point with pressure may be determined by enclosing the liquid to be boiled, reducing the air pressure on the surface and noting the temperature at which steady boiling takes place. This may be done by placing the flask under the receiver of an air pump. If the space over the liquid contains only the saturated vapor, and the pressure of this is suddenly reduced, as for example, by pouring cold water over a sealed flask containing

TABLE 12

CHANGE OF BOILING-POINT OF WATER WITH PRESSURE

Pressure (in mm. Hg.).	Boiling-point.	Pressure (in mm. Hg.).	Boiling-point.
680	96.91°C.	740	99.25
690	97.32	750	99.63
700	97.71	760	100
710	98.11	770	100.37
720	98.49	780	100.73
730	98.88		

water and its vapor slightly below the boiling point, boiling will at once begin. The pressure variation of the boiling-point of water is frequently used to determine the air pressure on mountain tops, and hence roughly their height. The variation for water is about 0.37°C. at 100° for a change in pressure of 1 cm. of mercury.

310. Other Conditions Affecting Boiling.—The ease with which vapor bubbles are formed in a liquid depends upon various conditions, such as the presence of dissolved gases, or of points or small solid particles in the liquid. Hence the prevention of "bumping," or the violent formation of vapor bubbles, is brought about by putting broken glass or other rough solids into the boiling liquid. The boiling-point of a given pure liquid is always raised by dissolving any relatively non-volatile substance in it, as, for example, sugar in water, but may be either raised or lowered by dissolving a volatile substance in it, as, for example, alcohol in water, which gives solutions boiling below the boiling-point of water, but above that of alcohol. With volatile combinations the boiling point may be either above or below the boiling point of *both* constituents, while with a non-volatile solute the elevation of the boiling point is, for dilute solutions, proportional to the mass of solute added and approximately the same for equal gram-molecular weights of all solutes in the same amount of a given solvent. For water the elevation is at the rate of 5° for each gram-molecular weight dissolved in 100 grams of water, and this is called the *molecular rise of the boiling-point.*

311. Specific Volumes and Densities of Vapors.—In general there is a very large increase in volume on vaporization. Thus 1 gram of saturated steam at 100°C. occupies 1721 c.c. while 1 gram of water at 100°C. occupies only 1.043 c.c. The volume of 1 gram of a substance is termed its specific volume, and is evidently equal to $\frac{1}{\rho}$, where ρ is the density. As the temperature is raised the specific volume of any saturated vapor decreases, whereas that of the liquid increases. Fig. 199 shows this behavior for water and steam. The two specific volumes of liquid and vapor become *equal* at a definite temperature, called *the critical temperature*, which will be further discussed in §316.

312. Heat of Vaporization.—The *heat of vaporization* is *the number of calories required to change* 1 *gram of a substance from a liquid at a certain temperature to a vapor at the same temperature under a specified pressure*, the symbol being L_t. The method of mixtures is usually used to determine the heat of vaporization, vapor from a boiling liquid being passed into a vessel immersed

in the water of a calorimeter, where the vapor is condensed and the heat given off. If M grams of vapor at a temperature t_3 in condensing raise m grams of water and the calorimeter (heat capacity $\Sigma m_1 s_1$) from t_1 to the final temperature t_2 of the mixture,

$$M[L_t + s(t_3 - t_2)] = (m + \Sigma m_1 s_1)\ (t_2 - t_1),$$

from which L_t can be computed, if s the specific heat of the condensed vapor, is known.

Since there is in general a very considerable increase in volume on vaporization, which occurs against a definite pressure, *external* work must be done by the vapor as it is formed. The heat of

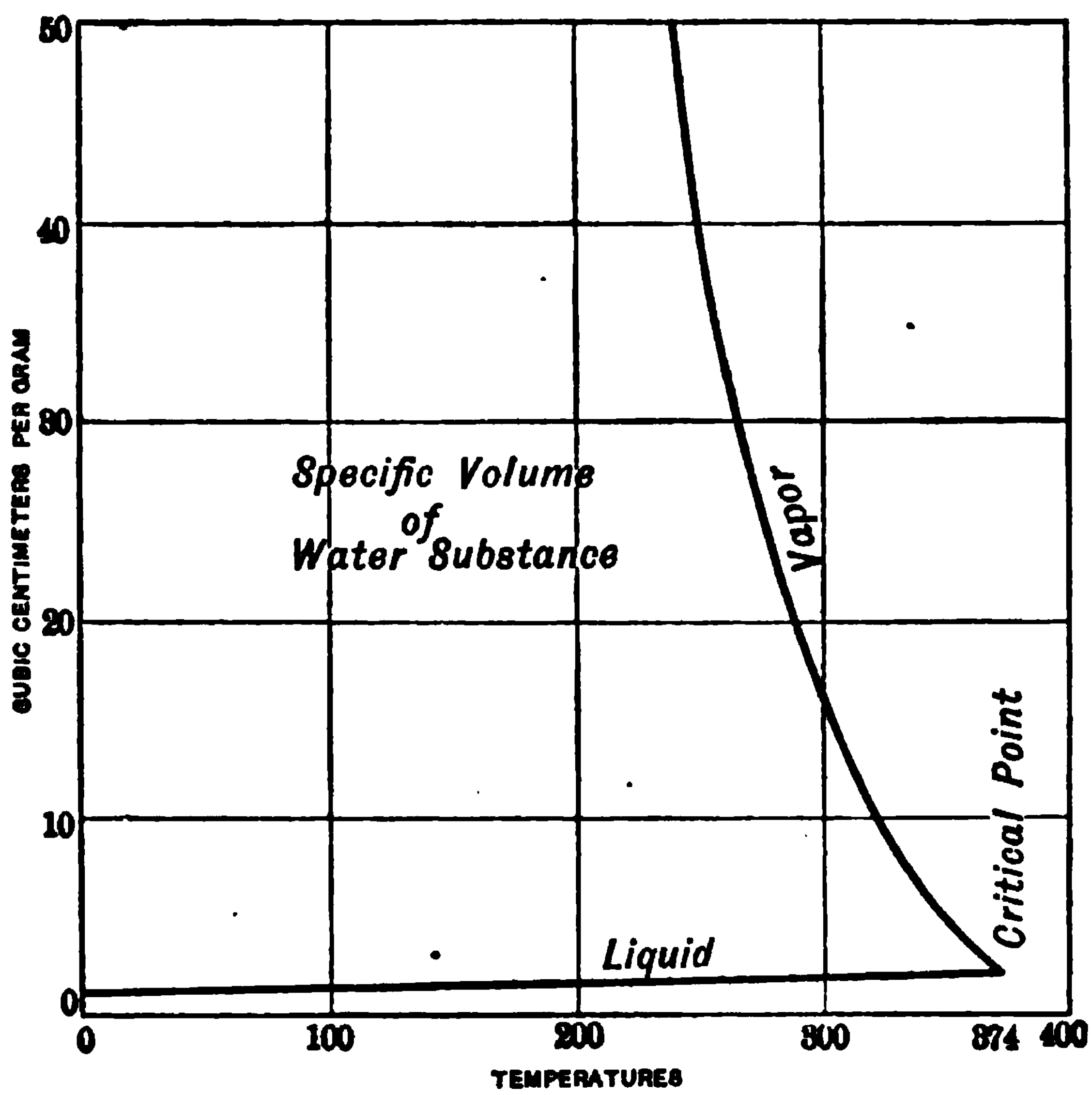


FIG. 199.—Specific volumes of liquid and saturated vapor at various temperatures.

vaporization may, therefore, be considered as made up of two parts, the *internal*, which is the increase in potential energy of molecules and atoms, and the *external*, which is the work done by expansion.

The heat of vaporization diminishes with increasing temperature and becomes zero at the critical temperature, where, as we shall see (§318), the distinction between liquid and vapor vanishes. The heat required to change 1 gram of a substance from

a liquid at 0°C. to a vapor at a temperature t is called the *total heat* of the vapor at temperature t. If H_t is the total heat, and we assume the specific heat is independent of temperature,

$$H_t = st + L_t$$

To supply this heat either the remaining liquid will be cooled, or heat will be drawn from the surroundings—hence the cooling effect of evaporation. The rate of evaporation will depend upon the rate of supply of heat; hence boiling over a fire will be more violent in a metal than in a non-conducting vessel. It is the heat of vaporization of steam which is chiefly effective in steam heating systems, this heat being supplied in the boiler and given up (potential converted to kinetic energy again) in the radiators where the steam is condensed. The use of a hand or electric fan in warm weather is the most common example of the reverse process, that is, cooling resulting from increased evaporation due to the circulation of the air.

TABLE 13

HEAT OF VAPORIZATION

(Calories per gram at normal boiling points.)

Alcohol (ethyl)	205	Liquid air	50
Liquid H_2	123	Liquid CO_2	96
Liquid O_2	58	Mercury	68
Liquid N_2	50	Water	542

313. Sublimation.—The direct change from the solid to the vapor state is termed *sublimation.* This or the reverse, direct condensation, is commonly observed in the evaporation of snow in cold dry weather, in the production of hoar frost, and in the evaporation and recondensation of camphor confined in a bottle. As in the other two transformations, there is a definite vapor pressure, for every temperature, at which the solid and vapor can exist in equilibrium together, without one state continuously changing into the other. When plotted on the Pt plane these points give the "hoar frost" or sublimation line, the equilibrium pressure falling with the temperature. Sublimation also involves an increase in potential energy and external work, and hence a *heat of sublimation.*

314. Unstable Conditions.—In stating that substances solidify, vaporize, and condense at *definite temperatures* under a given pressure, we have disregarded certain cases in which it is possible to cool a liquid very much below its freezing point without solidification and to heat a liquid very much above its boiling-point without vaporization. These, however, are abnormal and

unstable conditions, since, if freezing (or boiling) once starts, it goes on with great violence till the normal temperature has been restored. For example, small drops of water in an oil of equal density have been cooled at atmospheric pressure to −20°C without freezing and heated to 178°C. without boiling, and minute spheres of platinum and other metals have been cooled several hundred degrees below their normal melting point before solidification occurred. Agitation, the presence of points or particles, or the least trace of the solid serve to start solidification, while points and pieces of porous solids start boiling. A liquid above its boiling-point will begin to boil violently if touched by a file or paper—though these materials become ineffectual when they become *clean*.

Condensation of a vapor is difficult to start without the presence of *nuclei*, that is, dust particles, liquid droplets or electrified molecules or atoms, called *ions*, which very greatly assist the formation of large drops. The efficacy of some of these is due to their providing surfaces of relatively larger radius on which condensation can take place, since the vapor pressure necessary or the equilibrium of a vapor with a liquid drop is less for a drop of large radius, and also less for electrified drops.

315. The Triple Point Diagram.—Having discussed the three equilibrium curves, solid-liquid, liquid-vapor, and solid-vapor, let us consider them combined as in Fig. 200 in the *Pt* plane. Since the areas on either side of each cuver represent conditions such that only one state can exist, for example solid to the left and liquid to the right of the freezing-point curve, a little consideration will show that, in order to be consistent, the three curves must *intersect in a point*. If this were not so the area included between the curves would represent quite contradictory conditions as deduced from the several curves. Since each curve represents conditions of possible coexistence of two states, in the condition represented by the point of intersection all three, solid, liquid, and

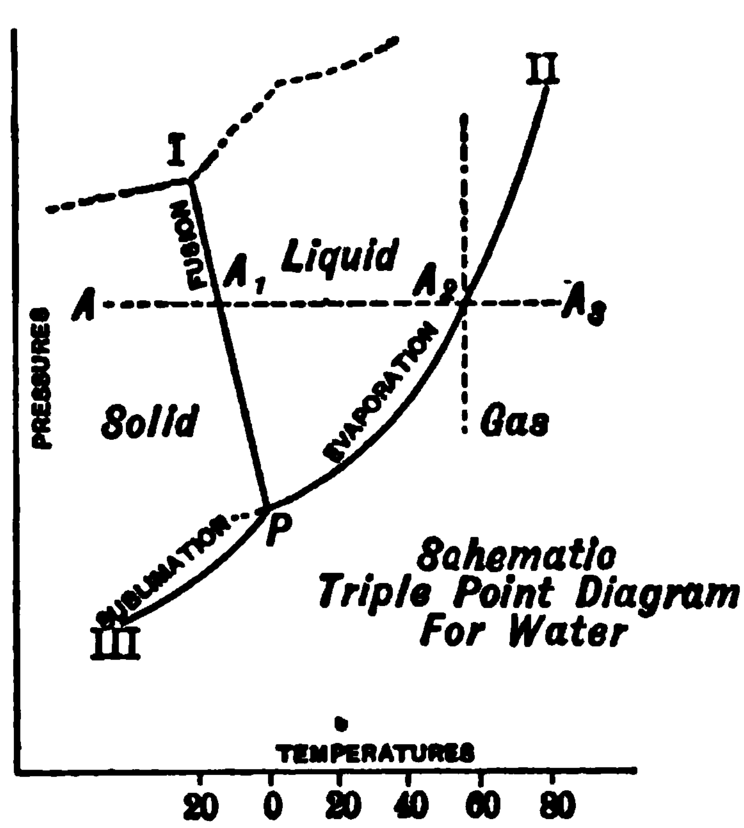


FIG. 200.—Triple point diagram.

vapor can exist simultaneously in equilibrium and this is the *only* condition in which this is possible. It is called the *triple point* and for water corresponds to +0.0072°C., and a pressure of about 4.6 mm. of mercury.

It may be thought that common experience contradicts this conclusion as to the unique properties of the triple point—ice, water, and water vapor being often found coexistent at various temperatures. Under these conditions it will be found, however, that the ice is melting and the vapor condensing, or water evaporating—at least *equilibrium does not exist*. The characteristics represented by the triple-point diagram are true for all substances.

316. The Critical Temperature.—In 1822 Cagniard de la Tour made important discoveries regarding the relation between the liquid and the vapor state by filling a glass tube with alcohol and its vapor, sealing off the tube and heating it. If about two-thirds of the total volume were liquid at ordinary temperatures he found that, as the temperature increased, the meniscus, or curved surface separating liquid from vapor and due to molecular attractions (§208), became flatter and less distinct and finally disappeared. Thus above a temperature of about 243°C. the liquid and vapor have the same molecular attractions (no meniscus) and are visually identical.

The limiting temperature at which a separation can be observed between the liquid and the vapor state is called the critical temperature. We have already seen that the specific volumes of a liquid and its vapor become more nearly equal as the temperature is raised, and that at the same time the latent heat of vaporization becomes less, and the relation of these facts will now be clear.

317. Isothermal Curves for CO_2.—Andrews, in 1863, inclosed CO_2 in a glass tube, kept this at a constant temperature and compressed the gas by forcing mercury into the tube. He measured the pressure required to compress the gas to various measured volumes while the temperature was kept constant, and did this at a series of temperatures from 13° to 48°C. Corresponding pressures and volumes plotted on the PV plane gave what we have called *isothermal curves*, shown in Fig. 201.

The effect of compression at 21.5° from an initial volume of 12 c.c. per gram is, as shown by the curve, first a gradual increase in pressure until a pressure of 59 atmospheres is reached (at A)

when liquid CO_2 will suddenly appear in the tube, after which no increase in pressure will occur in spite of diminution in volume until *B* is reached. During this time condensation has continued, until (at *B*) the vapor has been entirely changed to liquid, after

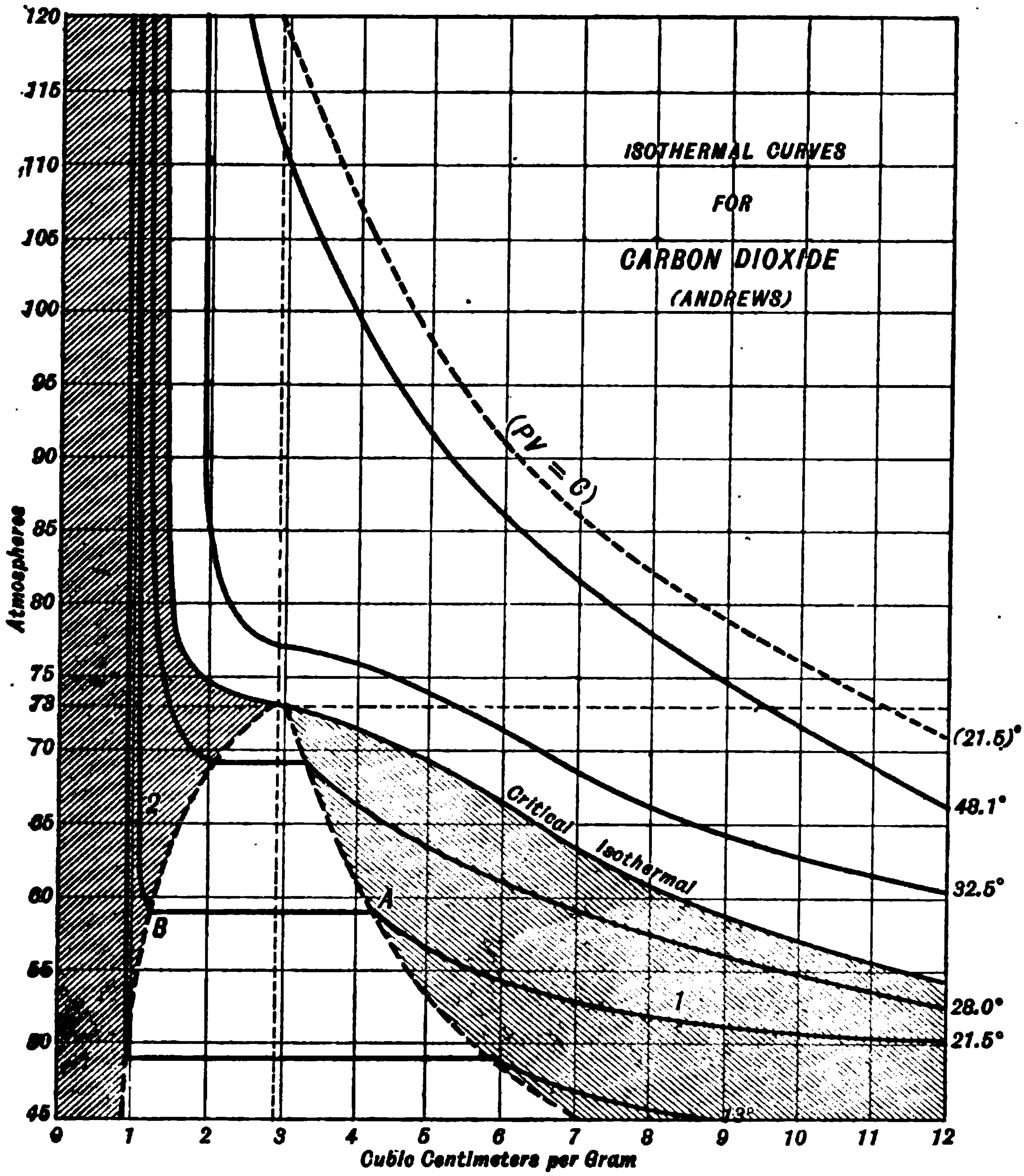


FIG. 201.—Isothermals of carbon dioxide.

which any decrease in volume necessitates a very great increase in pressure, the liquid being in general very incompressible.

If the same sequence of operations is carried out at a higher temperature, it is found that condensation begins at a less volume

and higher pressure, and ends at a greater volume, that is, the *volume interval* during which the substance is part vapor and part liquid with a visible meniscus between becomes less as higher temperatures are chosen, until, when a temperature of 31° is reached, the horizontal part of the isothermal has disappeared, and no separation into liquid and vapor can be noticed during compression. The critical temperature is 30.92°C., and it may be defined in another way as *that temperature above which it is impossible to liquefy a gas by pressure alone*, "to liquefy" meaning to cause a separation of the two states.

The line AB represents those conditions of pressure and volume in which liquid and vapor can exist in equilibrium with each other at the temperature of 21.5°, and a similar meaning attaches to the horizontal portions of the other isothermals. The dotted line through the ends of these horizontal lines surrounds an area representing all the physical conditions under which the liquid and its vapor can be in equilibrium with each other, and the highest point of this curve is the critical point whose coordinates are the *critical volume* and *critical pressure*, V_c and P_c, corresponding to the critical temperature t_c.

318. The Saturation Curve.—That part of the dotted curve to the right of the critical point is called the "saturation curve" and evidently represents all possible conditions of saturated vapor, and since the diagram is drawn for unit mass, the abscissæ of the points of this curve are the specific volumes of the saturated vapor at various temperatures. The left branch of the dotted curve is called the "liquid curve" and the abscissæ of this portion are the specific volumes of the liquid at the same temperatures. It is very clear, then, that the specific volumes of liquid and saturated vapor become equal at the critical point.

Above the critical point the distinction between liquid and vapor disappears, and the substance passes continuously and homogeneously from a rare, easily compressible condition which we would call gaseous, to a dense, almost incompressible condition, which we would naturally call liquid. It is possible, by properly varying the pressure, volume and temperature, to pass from any condition 1 to any condition 2 without crossing the dotted curve, that is, without having the liquid distinct from the

vapor at any time. This property is called the "continuity of state."

It is generally agreed to call a substance a vapor if its condition is represented by any point below the critical isothermal and to the right of the saturation curve, and a gas if represented by a point above the critical isothermal, though this distinction is not important. The properties represented by this set of isothermal curves for CO_2 are characteristic of all substances which have been studied.

TABLE 14

CRITICAL DATA

Substance.	Critical temperature °C.	Critical pressure (Atmos.).
Air	−140	39
Alcohol (ethyl)	243	62.7
Ammonia	130	115
Argon	−117	52.9
Carbon dioxide	30.92	73
Chlorine	146	93.5
Helium	−268.5	2.3
Hydrochloric acid	52.3	86
Hydrogen	−234.5	20
Nitrogen	−146	33
Oxygen	−118	50
Radium emanation	104.5	62.5
Water	374	194.6

319. Equations of State.—Many attempts have been made to derive equations for the isothermal curves of Fig. 201, corresponding to the equation $PV = RT$, which holds approximately for conditions far removed from the critical, but none have been entirely successful. One of the most satisfactory of such "equations of state" is that of Van der Waals:

$$\left(P+\frac{a}{V^2}\right)(V-b)=RT$$

in which a, b, and R are constants for a given substance. This agrees fairly well with the results of experiment, though, instead of the straight portion AB of the isothermal, the equation gives a continuous curve which cuts the straight line in three points as shown in Fig. 202. The possibility of a continuous passage such as $DCBA$, below the critical temperature, from the vapor to the liquid condition, was suggested by James Thomson shortly after

Andrews' work; but, except for portions of the curve from A toward B (under-cooling a vapor free from nuclei) and from D toward C (superheating a liquid), it has not been realized experimentally and indeed seems quite unrealizable, since it would represent states in which an increase in volume would accompany an increase in pressure.

Corresponding States.—It was suggested by Van der Waals that if the pressure, volume, and temperature were expressed in terms of the critical constants, P_c, V_c, t_c, for each substance, as units, instead of in atmospheres, cubic centimeters and degrees centigrade, for example, the "equation of state" would be the same for all substances, containing no constants peculiar to any one material. The states of all substances would *correspond* when they were represented by the same "reduced" values of P, V, and t. While this "theorem of corresponding states" is a necessary consequence of Van der Waals' equation, and may be safely and usefully applied between related substances, it is not in general true.

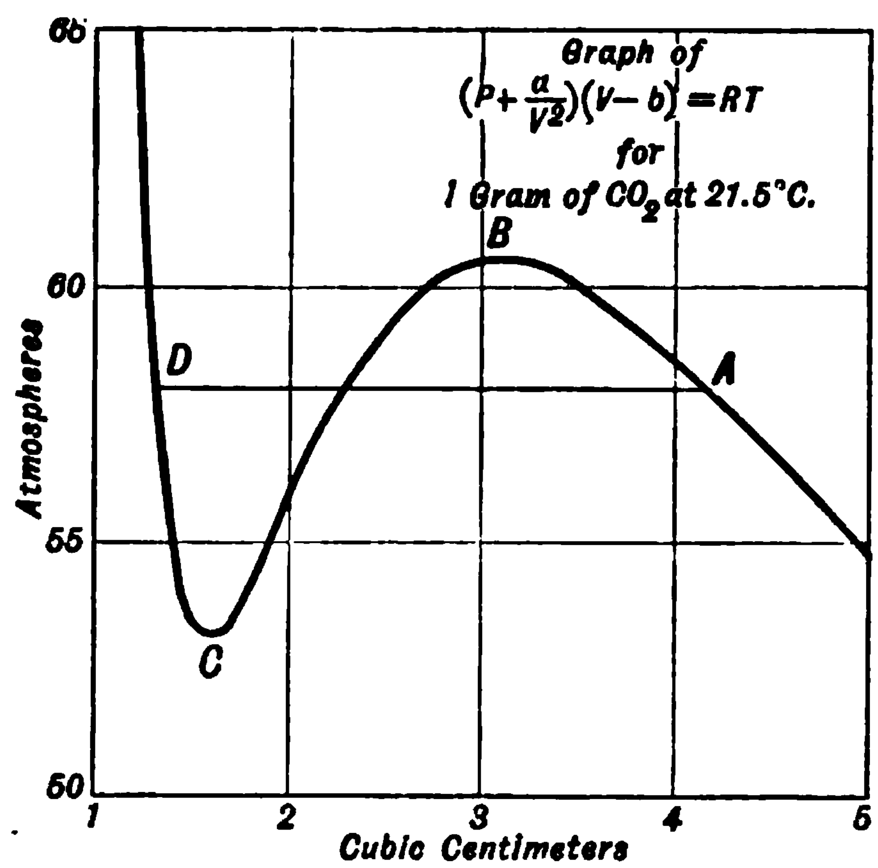


FIG. 202.—Graph of Van der Waal's equation.

320. Thermodynamic Surface.—If the isothermals of Fig. 201 are placed in their proper position along the temperature axis, a smooth surface drawn through them forms the thermodynamic surface shown in Fig. 203, every point of which represents by its coordinates P, V, t, an equilibrium condition of 1 gram of a substance. The conditions under which the substance may exist as liquid or gas or as a mixture of different states, are indicated on the diagram, and the triple point curve and the isothermals which we have already discussed are seen to be the projection on the Pt or Pv plane of lines on this thermodynamic surface.

321. The Liquefaction of Gases.—By compression and cooling Faraday (beginning in 1823) liquefied carbon dioxide, sulphur dioxide, chlorine and several other gases not previously known

in the liquid state. The temperatures he used were evidently below the critical temperatures as we now know them, but the problem was not thoroughly understood until the work of Andrews made it probable that extremely low temperatures as well as high pressures would be needed to liquefy oxygen, nitrogen, hydrogen and air, which, as late as 1877, were called *permanent* gases. The problem has been, then, one of devising methods of obtaining extremely low temperatures, and it has been so successfully solved that all known gases have now been liquefied. The following methods are used for obtaining low temperatures:

1. *Chemical method*, or the use of freezing mixtures, that is, mixtures of substances which in dissolving or combining absorb

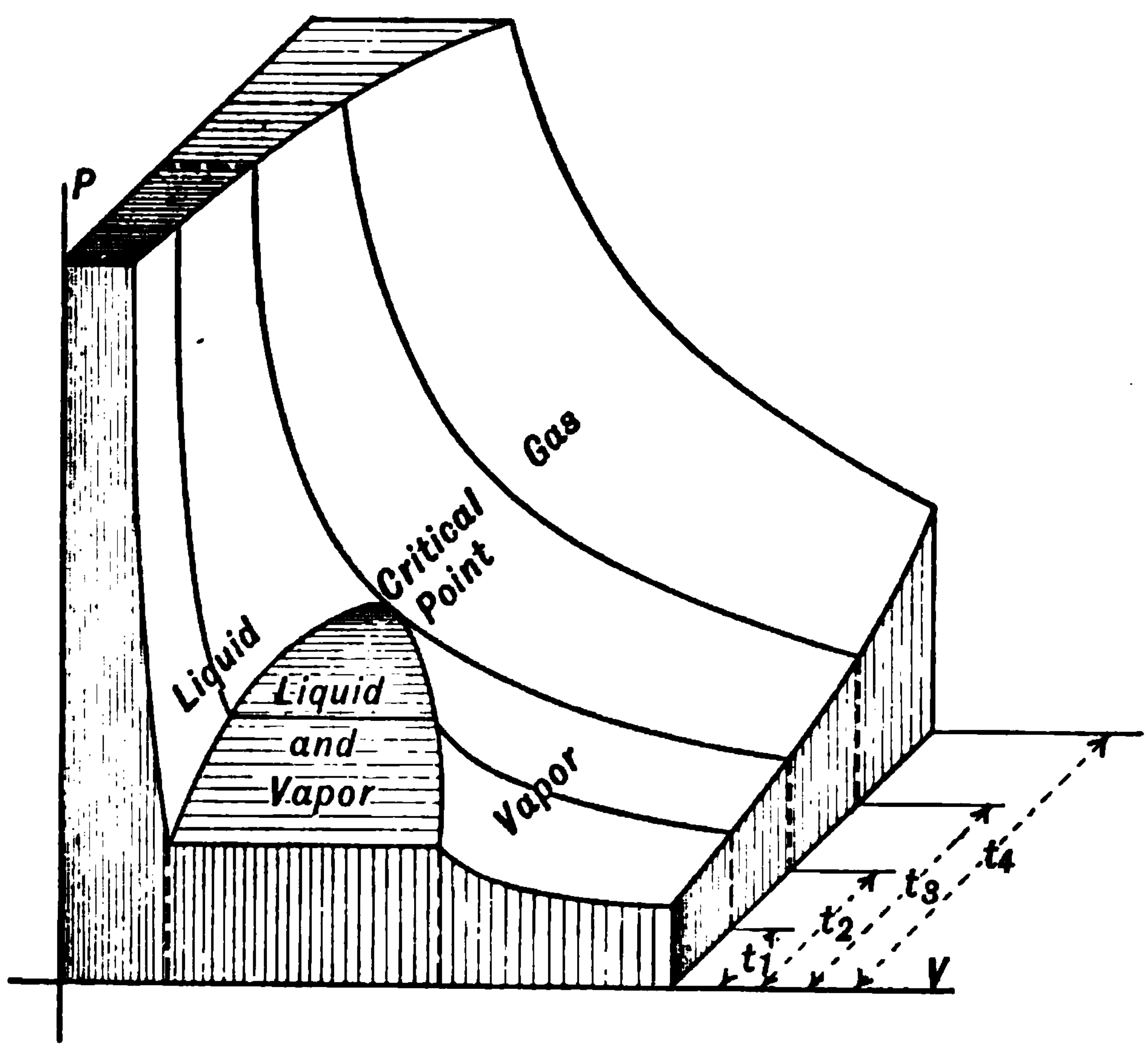


Fig. 203.—Thermodynamic surface, coördinates P, V, t.

heat and thus lower the temperature. The lowest temperature reached by this method, −82°C., has been obtained by mixing solid CO_2 and liquid SO_2. This method is not used in recent liquefying processes.

2. *Evaporation Method.*—Fig. 204 illustrates this method as applied to the ammonia refrigeration process. A compressor exhausts ammonia gas from above the liquid, compresses it, forces it through tubes cooled by running water where the heat of

vaporization and of compression is taken out and it again becomes liquid, and back through a reducing valve into the evaporating chamber. The evaporating chamber is surrounded by the material to be cooled (circulating brine in the ordinary refrigerating system), from which the heat necessary to vaporize the ammonia is absorbed. The ammonia therefore absorbs heat in the evaporation chamber and loses heat in the cooling coils.

A series of such circulating systems, containing, for example, SO_2 in the first and CO_2 in the second, arranged so that the cooling of the CO_2 is done by the evaporating of the SO_2 is called the *cascade method*, by which Pictet in 1877 liquefied oxygen. The oxygen was compressed to several hundred atmospheres pressure in a tube surrounded by the evaporating CO_2 and thus cooled to $-140°C$. Reference to Table 14 shows that at this temperature and pressure the oxygen would be liquid. Upon opening a cock the oxygen escaped in a white stream, indicating the presence of the liquid or solid. By adding lower steps to the cascade it is possible to obtain very much lower temperatures, so that oxygen, nitrogen and hydrogen may be obtained liquid at atmospheric pressure.

Fig. 204.—Ammonia refrigerating process.

3. *Regenerative Unbalanced-expansion Method.*—We have already considered (§296) the cooling experienced by all gases except hydrogen when forced through a small opening. This cooling is for air only about 0.25°C. per atmosphere decrease of pressure, but it increases with decreasing temperature. Hence if the cooled expanded gas is led back around the in-flowing compressed gas, as in Fig. 205, so as to cool it, the temperature at which expansion takes place will be gradually lowered, until finally some of the gas will liquefy on expansion.

Apparatus for applying this principle was independently invented by Linde, Hampson and Tripier, and this method is now extensively used,

commercially as well as scientifically, for liquefying oxygen, nitrogen and hydrogen. The expenditure in the work of compression of 10 h.p. for one hour will, by this process, produce from 2 to 4 liters of liquid air, the initial pressure being from 120 to 200 atmospheres. The heating which is observed with hydrogen at ordinary temperatures becomes zero at $-80.5°C.$, as shown by Olszewski, and below that there is a cooling, so that, by initially cooling hydrogen below $-80.5°C.$, it can be liquefied by the unbalanced expansion process, as was first done by Dewar in 1898.

4. *Regenerative Balanced-expansion Method.*—The first step in this method is the compression of the gas to about 40 atmospheres pressure, and the partial cooling of it by an "interchanger" analagous to the one used in the Linde process. The compressed and cooled gas is then admitted to a cylinder and allowed to expand against a piston thus doing external and internal work,

FIG. 205.—Linde's apparatus for liquefying air.

and being still further cooled to $-160°C.$ or lower. The expanded gas is then led around the outside of a liquefying vessel containing air at 40 atmospheres pressure, and cools sufficiently to liquefy some of the air. The expanded gas, after absorbing heat from the liquefying vessel, is led back through the interchanger to the compressor.

The advantage of this method over those of the Linde type lies in the greater amount of *external work* which the gas does, resulting in greater cooling. Moreover, being done against a piston, this work can be utilized. By combining three stages of expansion similar to the above, Claude has produced liquid air at the rate of 9 liters per 10 h.p. per hour.

In 1907 helium, the last gas to resist liquefaction, was liquefied by Kammerlingh Onnes by the unbalanced-expansion method, its boiling-point under one atmosphere pressure being −268.5°C. The lowest temperature attained by evaporating helium under a pressure of 1 cm. of mercury was −270°C. or 3° above absolute zero. These extreme temperatures are measured either by a gas thermometer containing helium at reduced pressure, or by a thermo-electric or resistance thermometer.

TABLE 15

BOILING-POINTS OF DIFFERENT SUBSTANCES UNDER ATMOSPHERIC PRESSURE, AND TEMPERATURES OBTAINED BY BOILING UNDER REDUCED PRESSURE

Substance.	Boiling-point (Atm. pressure).	Pressure (in mm. of mercury).	Boiling-point (reduced pressure).
Argon	−186.2	300	−194.2°C.
Carbon dioxide	− 78.2	2.5	−130
Helium	−268.5	10	−270
Hydrogen	−252	....	−256
Neon	−109	2.4	−257.5
Nitrogen	−195.8	86	−210.6
Oxygen	−163	200	−194
Radium emanation	−62	9	−127

TRANSFER OF HEAT

322. Convection, Conduction and Radiation.—Heat is transferred by three very different processes.

Convection is the transport of heat by moving matter, as, for example, by the hot air which can be felt rising from a hot stove.

Conduction is the flow of heat through and by means of matter unaccompanied by any motion of the matter, for example, the passage of heat along an iron bar one end of which is held in a fire.

Radiation is the passage of heat through space without the necessary presence of matter, for example, the passage of heat through the vacuum in the bulb of an incandescent lamp.

323. Convection.—Convection occurs in liquids and gases and is due to the change in density produced by rise in temperature.

A volume of liquid or gas which varies in density in different parts is only in stable equilibrium when the densest portions are at the bottom, and there is a regular decrease in density towards the top. Since (with the exception of water below 4°C., §278), liquids and gases expand on heating, thus diminishing in density, the heated portion will *rise* and there will be an upward convection current of hot substance and a downward convection current of cold substance to take its place. If heat is added at the *top* of an enclosed liquid or gas, there will be no convection, (except with water below 4°C.).

Common examples of *convection by liquids* are the distribution of heat through liquids heated from the bottom, as in the case of water in a tea kettle, and the distribution of heat through a house by the hot water system of heating, Fig. 206. The water is heated in A, rises through B, is cooled in the radiator and falls through C. On a large scale the Gulf Stream, Japan current, and other warm surface ocean currents which start near the equator are, in part at least, caused by convection, the return being a cold current flowing toward the equator along the ocean bed.

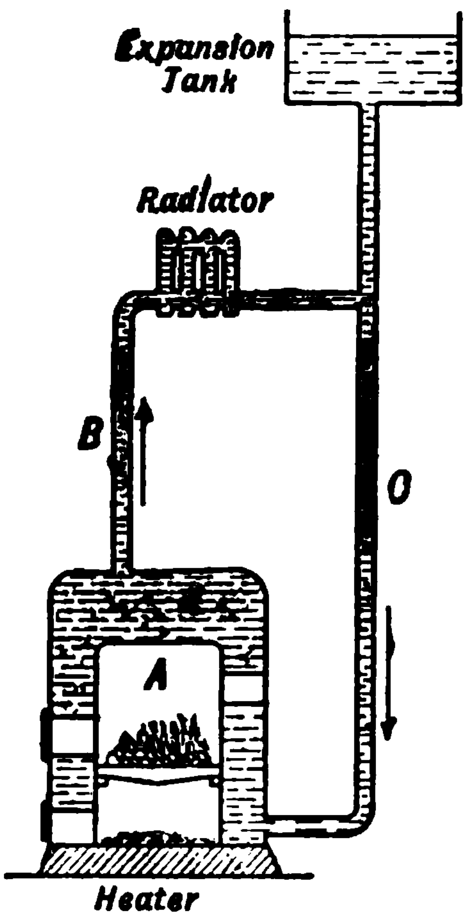


FIG. 206.—Transfer of heat by convection of hot water.

The hot-air furnace system of heating houses is based on *convection by gases*, the hot air rising from the furnace through the pipes and registers, and the supply of cold air coming usually from outside. The working of such a system can sometimes be improved by establishing a direct return from the coldest part of the house to the furnace, thus completing the indoor circulation.

"Natural ventilation" is also a convection process, an outlet being provided at the top of a room for the warm stale air, and an inlet at the bottom for the cool fresh air. The natural draft in chimneys has a similar cause; the higher the chimney the larger is the undisturbed column of warm air and hence the greater the draft. The mixing of currents of hot and cold air usually cause a flickering or "boiling" of objects seen through them, because light travels differently in hot and cold air. This effect may be seen by looking across a flat country in the hot sunshine, or over a hot pavement or stove.

The winds are largely convection effects, the simplest example being the "land" and "sea" breezes, which ordinarily blow from the sea to the land in the morning and from the land to the sea at night. These are the return currents which replace warm air which rises from the quickly heated land in the morning and from the warmer, more slowly cooling sea at night.

CONDUCTION OF HEAT

324. Characteristics of Heat Conduction.—As we have said, conduction is the flow or passage of heat energy through and by means of matter unaccompanied by any obvious motion of matter, as, for example, the passage of heat through the bottom of a kettle to the water inside.

The direction in which heat will flow between two points, whether from A to B or B to A, is found to depend on the relative temperatures of A and B, heat always flowing from the point at higher temperature to the one at lower temperature. The greater the difference of temperature beween two points, other conditions being the same, the more heat will flow per second, for which reason a kettle boils more quickly over a hot fire than over a low one. But, with a given temperature difference between the fire and the water, boiling will take place more quickly with a thin-bottomed kettle than with a thick. The temperature difference between two points, A and B, divided by the distance, l, between them, or $\frac{t_A - t_B}{l}$, which is the average fall in temperature per centimeter between A and B, is called the *temperature gradient*. The above statements of the dependence of conduction on temperature difference and distance may be combined by saying that the amount of heat conducted per second between two points is *directly proportional to the temperature gradient.* To pursue the same example, common experience dictates that the kettle should have a broad bottom in contact with the stove. This is an illustration of the fact that, other conditions being the same, the amount of heat conducted per second is *directly proportional to the area* through which it can flow.

Finally, the rate of flow of heat, other conditions being the same, depends greatly upon the material through which it must

flow, substances being roughly divisible into "good conductors," which permit under given conditions a large flow of heat, and which in general are metallic, and "poor conductors," which permit a small flow of heat and are in general non-metallic, such as wood, glass, asbestos, leather, linen. Examples of this difference are very common. A glass of hot water may be handled, while a metal cup containing the same water will be too hot to touch. Handles of heating vessels are made of wood, or, if of metal, are covered with string or cloth, so that they may be touched. Given two bodies, one metal and one wood, at the same temperature, below that of the hand, the metal one feels much cooler because the heat it takes from the hand quickly spreads through the mass, while, with the poor conducting wood, the heat remains near the surface of contact, which quickly rises in temperature. Thus the wood feels warmer because (after the first instant) it *is* warmer, where it is touched. The rate of heat flow through a body will depend not only on the substance composing it but upon its condition of subdivision and density. Thus saw-dust conducts less readily than wood, and the small conduction through cork is partly due to the reduction of effective cross-section by air holes. Also substances when moist conduct better than when dry, because water, which fills the pores, is a better conductor than air.

The characteristic of bodies which determines the rate of flow of heat through them is called their *thermal conductivity*, and the numerical measure of this characteristic, is called the *coefficient of thermal conductivity* which will be defined in the next paragraph.

325. The Coefficient of Thermal Conductivity.—In order to group together the previous statements and obtain an exact definition of the conductivity coefficient, consider the passage of heat from the region (2), Fig. 207, at the uniform constant temperature t_2, to (1) at the temperature t_1, by conduction along the rectangular bar of cross-section A $(=ab)$ and length l, no heat being allowed to escape from the sides of the bar. If H is the heat which passes in a time T, then from the statements above, it follows that, for a given substance,

$$H \propto \frac{T(t_2-t_1)A}{l} \tag{1}$$

and, by introducing a proportionality factor K, properly chosen for each substance, this may be written as an equality,

$$H = \frac{KT(t_2 - t_1)A}{l} \tag{2}$$

K is the coefficient of thermal conductivity of the material. If, as a special case, we take $A = 1$, $T = 1$, $t_2 - t_1 = 1°$, $l = 1$, then $K = H$, or K for a given substance is *the heat flow per unit time per unit area with unit uniform temperature-gradient.* In the c.g.s. system, the units would of course be the centimeter, second, centigrade-degree and calorie. This statement defines K at a definite mean temperature, $(t_1 + \frac{1}{2})°$, but, since K varies with the temperature, equation (2) will not be true for any great difference of temperature, $t_2 - t_1$, unless K stands for the *mean value* between these limits, and the temperature gradient is *uniform.*

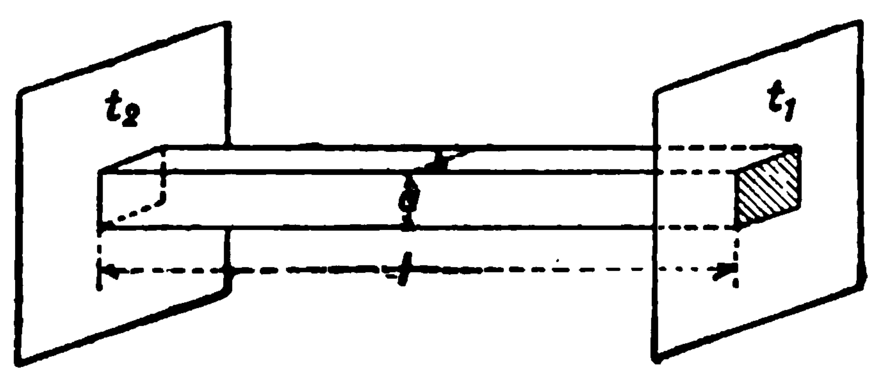


FIG. 207.—Illustrating simple case of conduction of heat.

Equation (2) is the basis of most methods for measuring K, but while they are very simple in principle they are very difficult to carry out accurately.

If a uniform bar of the substance has its end faces maintained at fixed temperatures, for instance $t_2 = 100°$C., and $t_1 = 0°$C., by contact with steam and melting ice respectively, its sides being protected from loss of heat, and if we find, by measuring the amount of ice melted, the heat which flows through the bar in a time T, and also the length and cross-section of the bar, K may be computed at once. This will, of course, be the mean value of K between 0° and 100°C. Other calorimetric methods are usually used for measuring the amount of heat flowing in a given time, and with poor conductors a slab or plate of the substance rather than a rod must be used to give a measurable rate of flow.

The average temperature gradient in the earth's crust is about 1°C. rise per 144 ft. of descent near the surface, increasing to 1° in 90 ft. at depths of a few thousand feet. This means a continual loss of heat from the interior, small in amount, however, on account of the low conductivity of rocks and soil. For the same reason (low conductivity) the daily variations of surface temperature penetrate only about 3 ft., the annual about 50 ft. (§238).

326. Conduction in Liquids.—In order to measure conduction in liquids, as distinct from convection, heat must, of course, be added at the *top* (except with water between 0° and 4°C.), since, if heated from below, the warm expanded liquid would rise, while, if heated from above, it will remain in place. Since many liquids are transparent to radiation (§330) there is also danger that we may confuse radiation across the liquid with conduction through it. The conductivity of liquids is, in general, about that of solids of low conductivity, except in the cases of mercury, which is metallic and a good conductor, and water and some aqueous salt solutions, which are intermediate between the metallic and the non-metallic solid conductors.

327. Conduction in Gases.—The masking of conduction by convection and radiation is even more likely to occur in gases, because of their greater mobility, greater transparency, and lower *real* conductivity. The conductivity of hydrogen and helium is much greater than that of other gases. This follows naturally from the kinetic theory, since they have the smallest molecular weights, therefore the highest molecular velocities at a given temperature, and therefore hand on kinetic energy from molecule to molecule with the greatest rapidity. The conductivity of gases is, through wide ranges, independent of the pressure as theory also indicates should be the case. On account of their extremely low conductivity air layers enclosed in or between solids, such as air spaces in house walls and in the walls of refrigerators, in pores in cloth and in fur or feathers, are chiefly responsible for the low conductivity found in these cases.

328. Conductivity of Alloys and Crystals.—The conductivity of copper is increased by compression, that of steel diminished by hardening. The conductivity of alloys is not, in general, simply proportional to the relative amounts of the pure metals forming the alloy, but may have decided minimum values in case compounds are formed. In non-isotropic solids, such as wood and crystals, the conductivity depends upon the direction of flow, being in the case of wood two or three times as great along the fiber as at

right angles to it. In crystals the axes of symmetry for heat conduction coincide with the crystalline axes, and the conductivity is different in different directions, as may be very prettily shown by means of a thin plate of crystal coated with wax on one side, and having a wire passing normally through the center. If the wire is carefully heated the wax will gradually melt and the limit of melting will be, in general, an ellipse and not a circle, as it would be with an isotropic plate. The marked decrease in effective conductivity resulting from breaking up a solid has been referred to, but this, as well as the effect of compression on a substance like felt or cotton, is not a change in the property of the substance itself, but merely a change in the amount of poorly conducting material (air) mixed with it.

TABLE 16

THERMAL CONDUCTIVITIES

(c.g.s. units.)

Substance.	Conductivity.	Substance.	Conductivity.
Aluminum	0.504	Lead	0.083
Brass	0.260	Nickel	0.142
Air	0.00005	Oak	0.0006
Concrete	0.0022	Platinum	0.166
Copper	0.918	Porcelain (Berlin)	0.0025
Cork	0.00013	Quartz, $\parallel$-axis	0.030
Cotton wool	0.00004	Quartz, $\perp$-axis	0.016
Earth's crust	0.004	Sawdust	0.00012
Flannel	0.00023	Silk	0.00022
Glass	0.0024	Silver	0.974
Gold	0.700	Tin	0.155
Ice	0.005	Water	0.0014
Iron	0.144	Zinc	0.265

329. The Nature of Conduction.—Since we have agreed that heat energy is in large part kinetic energy of motion of molecules, atoms, and electrons, it is natural to think of this motion (heat) as spreading through a substance by collision of these particles with each other. Heat added to one side of a body will increase the average energy of motion of the molecules on that side, and will be gradually handed on by impact to the slower moving ones and so will spread through the mass, much as a disturbance originating at one point would spread through a closely packed crowd of people by repeated pushing and jostling.

While this has been the common idea of the nature of the process, J. J. Thomson and others have recently attempted to account for the matter in an entirely different way, namely, by the *convection* of "free" electrons as

defined in §262. According to this hypothesis the addition of heat to one part of a substance increases the kinetic energy of the free electrons in this region, and there results, not only the transfer of energy by impact, but the actual diffusion of fast moving electrons from the hot to the cold region. The motion of these electrons, according to this hypothesis, also constitutes an electric current, so that this new explanation of heat conductivity would very easily account for the remarkable observed fact that those substances which conduct heat readily, such as metals, also conduct electricity readily, the electrical and thermal conductivities being in a fairly constant ratio for most metals at ordinary temperatures.

RADIATION

330. Radiant Energy.—Radiation is the process by which energy is transmitted through space without the necessary presence of matter. While being transmitted in this way energy is called *radiant* energy, and is not heat, since the latter is energy in a particular relation to matter. That energy may pass through matter and still not be heat may be shown by allowing the sun's rays to pass through glass and fall upon a blackened thermometer, which may be very decidedly heated, though the glass remains cool. Radiation differs most strikingly from convection and conduction in *speed*. Time a convection current (by means of smoke or dust) and the velocity will usually not be many feet per second; thrust one end of a silver rod into hot water and it will be several seconds before a noticeable effect can be felt a few centimeters above the surface; but an opaque screen for cutting off radiation produces a practically instantaneous effect even at a great distance.

The early idea regarding the nature of radiation was that it was a streaming of fine particles—that is, a convection. It is now known to be a *wave disturbance*, such as has been discussed in Wave Motion, analogous to the waves which travel over a water surface. The "disturbance" of which the water waves consists is an up-and-down motion of the water particles, and this disturbance travels *forward* while the water moves up and down. The "disturbance" in a radiation wave is a transverse (§238) electric (and magnetic) force (§543), which changes in direction and amount as a wave passes a point (just as the motion of the water particles changes from up to down), and which would move a compass needle if we could make one small enough for

it to act on. The characteristics of a radiation wave are the *period*, the *wave-length*, and the magnitude of the electric force which a wave produces as it passes, or the *amplitude*, corresponding to the height of a "crest" in a water wave, which determines how strong the wave is, how much energy it represents (§259), and is quite independent of the wave-length. A strong wave may be long or short, a long wave weak or strong.

331. Light and Radiation.—Radiation travels with the same speed as light (§641), and like light it can be reflected by mirrors and refracted by lenses and prisms. These and other facts prove conclusively that radiation waves are of exactly the same nature as light waves, in fact that *light* consists simply of those radiation waves whose lengths lie between 0.0004 and 0.00076 millimeters and which affect the eye. These waves lie near one end of the entire known range of radiation wave-lengths, which is from 0.0001 mm. to 0.3480 mm. and is called the *radiation spectrum*. Of this the *visible spectrum* evidently forms but a very small part. It is sometimes convenient to represent the spectrum by points along the x-axis whose abscissæ are proportional to the wave-lengths. That part of the spectrum called the "infra-red," of longer wave-length than the visible, contains usually the waves of greater energy, the most important for the radiation of heat, and these waves were formerly called "radiant heat." Radiation waves can travel through *free space*, their transmission being one of the fundamental properties of space as we know it.

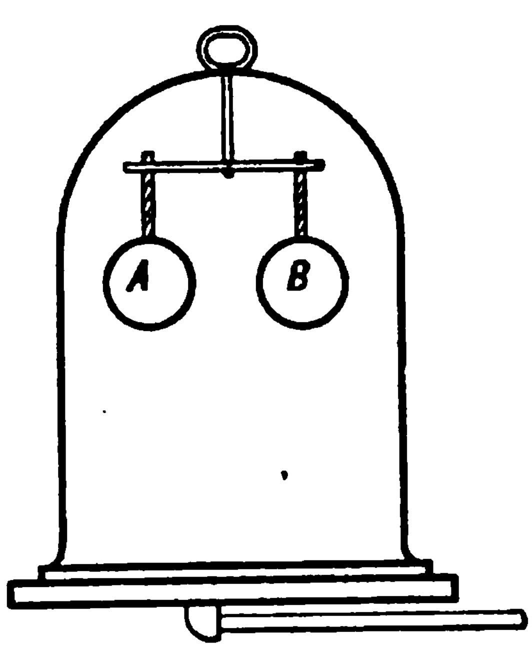


Fig. 208.—Radiating bodies in a vacuum.

332. Law of Exchanges.—If two bodies at different temperatures, for example the copper balls A and B, Fig 208,. are put in a vacuum and not in contact, equality of temperature will be established by radiation, the hotter body A on the whole radiating heat to the colder B. If, without changing the temperature or condition of B, A is cooled till it is the colder of the two, the net exchange of heat by radiation will now be from B to A. Since we have not altered B in any way, we conclude that B was

radiating to A in the first case also, but that A was radiating more to B. Radiation is then always a reciprocal process, or one of exchange. This is the celebrated *Prevost law of exchange*, according to which radiation equilibrium is the result of equal streams of radiant energy in opposite directions, and does not indicate the cessation of radiation.

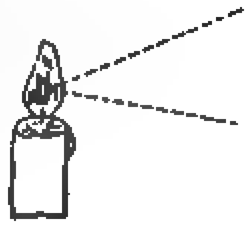


FIG. 209.—Thermopile for measuring radiant energy

333. Measurement of Radiation.—In order to measure radiation it is converted into heat by absorption in matter, the heat being then measured by the temperature change which it produces. To make this process delicate, recourse is had to the thermo-electric or resistance methods for measuring temperature described in §§268, 269.

Detail of Junctions

FIG. 210.—Detail showing arrangement of the exposed (inner) and protected (outer) junctions of a thermopile.

The *thermopile*, Figs. 209, 210, consists of one or more "junctions" of different metals, iron and constantan, or better two alloys of bismuth-antimony and antimony-cadmium arranged as shown in Fig. 210, so that one set of similar junctions can be exposed to radiation while the other set is protected. To increase the amount of radiant energy intercepted, the exposed junctions should be covered with light blackened silver or copper disks, and similar disks should be put on the other junctions. If the final elements are connected by wires to a very delicate galvanometer, very slight changes in temperature of one set of junctions, of the order of $\frac{1}{1,000,000}$°C. or less, will produce a readable deflection, and will correspond to a very weak stream of radiant energy falling on the exposed junctions, such as for example, the radiation from a single candle at a distance of 50 meters. To be quick-acting and sensitive the mass of the junctions should be small.

The bolometer, Fig. 211, is an even more sensitive instrument. It consists essentially of two similar strips of very thin (0.001 mm.) blackened platinum mounted side by side, having exactly the same resistance, and

arranged in a Wheatstone bridge (§456), so that any unequal changes in resistance of the strips can be very sensitively measured. If one strip is exposed to radiation, its temperature and hence its resistance will change.

334. Emission, Absorption and Reflection.—*Emission* is the starting of radiation waves. The *conversion of the energy of a wave into heat by passage through matter is called absorption.* A substance is opaque to radiation when it will not allow the radiation to pass through, as, for example, wood and metals are opaque to light. Absorption by a very thin surface layer of a strongly absorbing substance is called *surface absorption.* Such a surface, if polished, will also, in general, reflect very well.

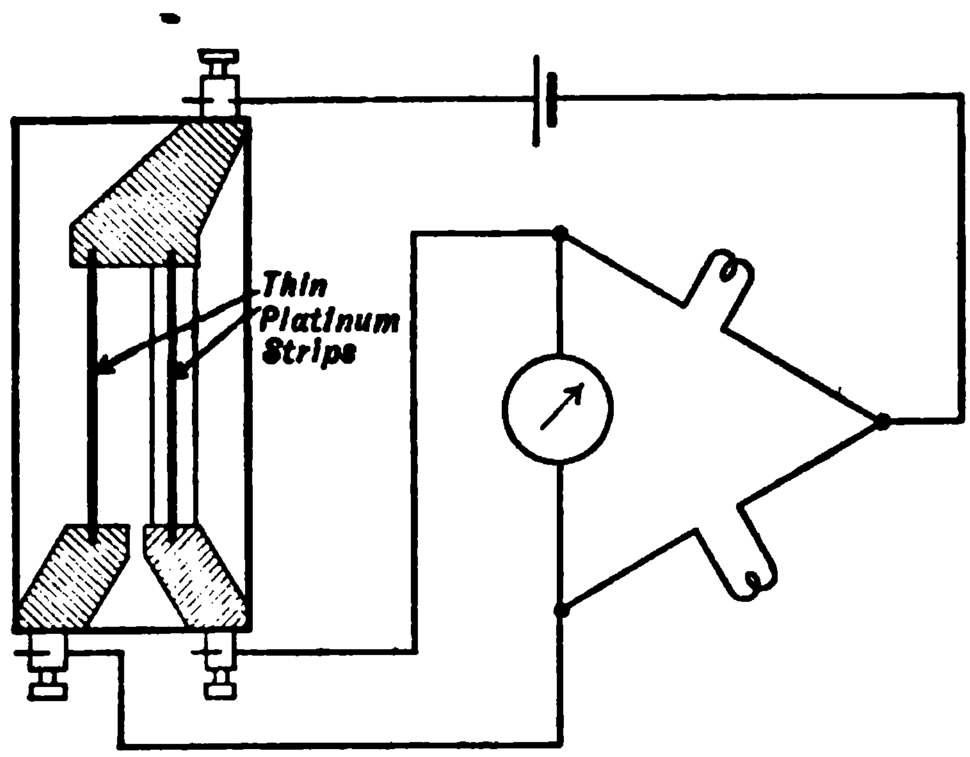


FIG. 211.—Bolometer for measuring radiant energy, and Wheatstone's bridge for measuring change of its resistance.

The *absorbing power* of a surface is the ratio of the radiant energy absorbed by the surface to the amount incident upon it. The absorbing power of a given surface is, in general, different or different wave-lengths of radiation. Let A_λ be the absorbing power for the wave-length λ and A the total absorbing power for all wave-lengths. The values of A and A_λ are practically independent of temperature. The *emissivity* of a surface is the total radiant energy, in ergs, which the surface sends out per square centimeter per second, this radiation being caused by the heat of the surface. The hotter a body the more it radiates, that is, emissivity increases with temperature. We shall denote the *total emissivity* for all wave-lengths by E, and the emissivity for the wave-length λ, or *partial emissivity, by* E_λ.

The *reflecting power* of a surface is the ratio of the radiant energy reflected from the surface to the energy incident upon it. The reflecting power of a surface is different for different substances and for different wave-lengths of radiation. Thus a polished silver surface will reflect about 82 per cent. of all blue light falling upon it, about 92 per cent. of incident yellow light, and about 98 per cent. of energy in the form of long infra-red waves, while a polished iron surface reflects about 57 per cent. of yellow light and from 78 to 97 per cent. of the energy of infra-red waves.

That there is a close connection between the absorbing power and emissivity of a surface can be shown, for example, by heating a bit of white china with blue markings, which look *dark* against the light china at ordinary temperatures because they *absorb* more light, but look *bright* against the china at high temperatures, showing that they *emit* more light. Similarly black ink marks on platinum look bright when heated. *In general good absorbers are good radiators.* That this must be so follows from a consideration of a body B suspended inside an exhausted opaque vessel C. Experience shows that B and C will come to the same temperature by interchange of radiation, and when equilibrium is reached B must absorb per second as much as it radiates. Hence if B is a good absorber it must be a good radiator, and vice versa.

335. Kirchhoff's Law.—The exact relation between absorbing power and emissivity. deduced theoretically by Kirchhoff and called after him *Kirchhoff's Law*, is that *the ratio of the emissivity to the absorbing power is the same for all surfaces at any one temperature*, or

$$\frac{E}{A}=\mathrm{E}$$

and similarly, as regards any particular wave-length,

$$\frac{E_\lambda}{A_\lambda}=\mathrm{E}_\lambda$$

where E and E_λ are constants independent of the substances.

336. A Perfect Absorber and Perfect Radiator.—If we could have a surface which *absorbed all the radiation alling upon it*, called a *perfect absorber* or *black body*, then for this surface

$$A=A_\lambda=1, \text{ and consequently } E=\mathrm{E} \text{ and } E_\lambda=\mathrm{E}_\lambda.$$

In other words, the constants E and E_λ are the total and partial emissivities of a black body. Since A and A_λ can never be greater than 1, it follows that a *black body* has the *greatest possible* total and partial *emissivity*,

at any temperature, and it is, therefore, also called a *perfect radiator*. A hollow opaque body having a *small* opening in the walls is a very close practical approximation to a black body, because radiation entering through the opening is partially reflected and re-reflected inside and thus eventually almost all absorbed. Also a sharp conical hollow or wedge-shaped cleft with straight opaque polished sides, no matter of what they are made, absorbs all radiation entering it. Conversely, if the walls of the enclosure, or cone, or cleft are uniformly heated, the radiation which leaves the opening will be that of a *perfect radiator* at the temperature of the walls, since, as we concluded in §335, it must be independent of the nature of the enclosure. These are all practicable ways of realising a *perfect absorber* and *perfect radiator*.

337. Total Radiation and Temperature.—The radiation of all bodies increases with the temperature, but the laws governing this increase are not as yet known except for a perfect radiator, for which Boltzmann deduced in 1883 the law previously suggested by Stefan, that

$$E = sT^4, \tag{1}$$

T being the absolute temperature of the surface and s a constant which later work has shown to be approximately 5.6×10^{-5} ergs per square centimeter per second. According to this law the radiation from one square centimeter of black body surface at 400° absolute (127°C.) would heat one gram of water 1.5°C. per minute. If one black body surface at temperature T is radiating to another surrounding it at temperature T_1, then the net or differential radiating power will be, from the law of exchanges,

$$E = s(T^4 - T_1^4)$$

While this law can be deduced only for a black body, it is found to hold approximately for other surfaces. Rewriting it in the form

$$E = s(T - T_1)(T^3 + T^2T_1 + TT_1^2 + T_1^3),$$

it is evident that, if T_1 does not differ much from T, we have approximately

$$E = 4sT^3(T - T_1) = K(T - T_1) \text{ (for } T \text{ constant)} \tag{2}$$

A similar relation, known as *Newton's law of cooling*, is found to hold for the loss of heat by *combined radiation* and *convection*, and was enunciated by Newton as follows:

The heat lost by radiation and convection by one body to another surrounding it is proportional to the temperature difference between

the two. This is a convenient relation to use and is quite accurate for small temperature differences.

338. Distribution of Energy in the Spectrum.—As the temperature of any radiating surface is raised, the energy emitted in every wave-length increases also, but not in equal proportion. It is a matter of common experience that the light emitted from a

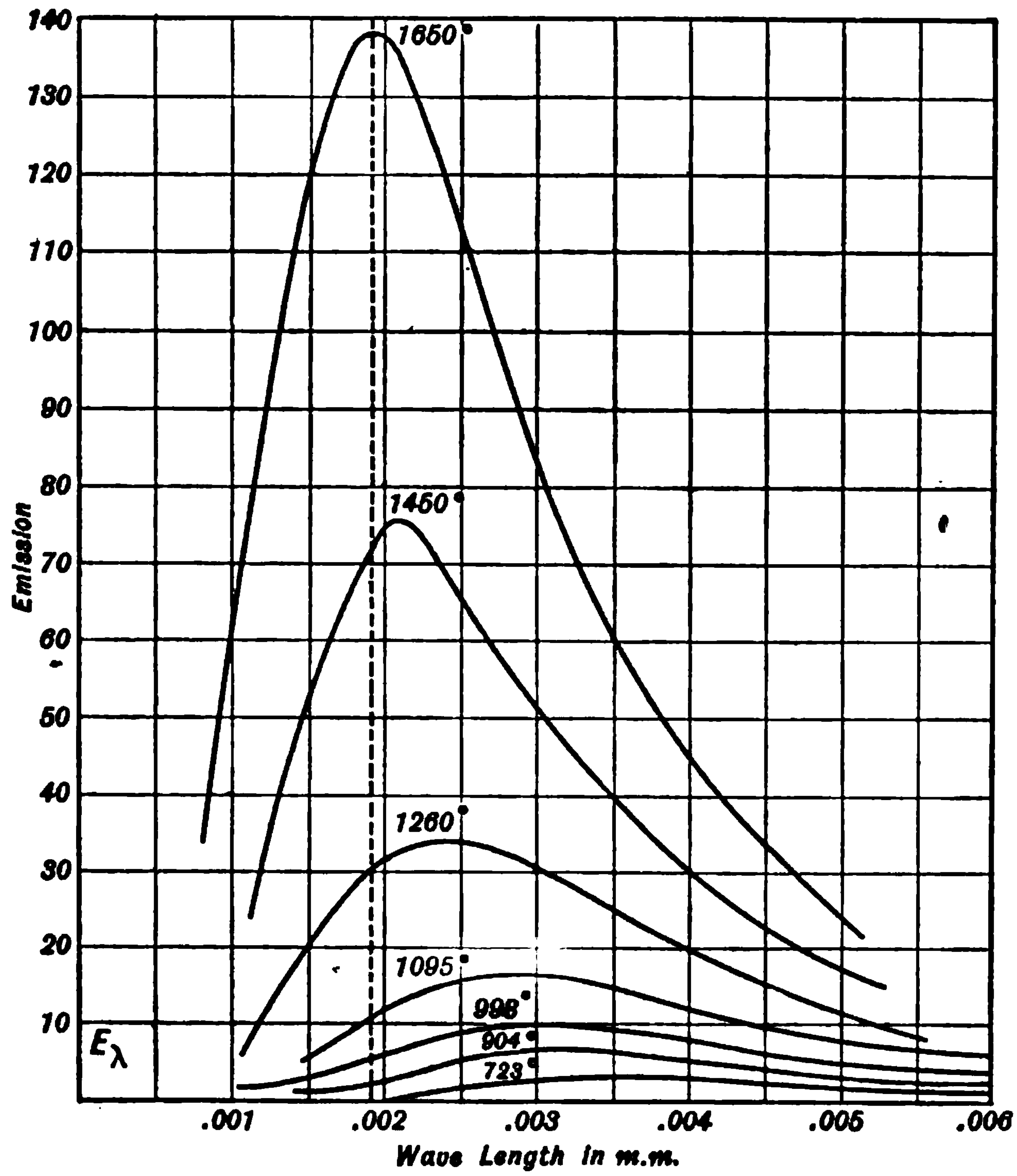


FIG. 212.—Curves showing distribution of energy in the spectrum of a perfect radiator at various temperatures.

hot radiating surface changes in color as the temperature of the surface is raised, changing from red to yellowish, then to white and finally having a blue-white color at extremely high temperatures.

If, for a given surface, we plot the values of E_λ as ordinates, and the corresponding values of λ as abscissæ, for any one temperature, we obtain what is called the *energy curve* for this temperature.

For example, the energy curves of Fig. 212 show clearly the distribution of energy in the spectrum of a perfect radiator at several temperatures. Such curves have a general similarity for all surfaces, the emission being weak for short wave-lengths, rising to a maximum, and diminishing again for long wave-lengths. As the temperature of the radiating surface is increased, all the ordinates of the curve increase, and the maximum shifts toward the short wave-lengths. This shift of the energy curve, resulting in an increasing proportion of blue in the emitted light, accounts for the change in color of an incandescent body, which was just referred to. For a perfect radiator at 100°C. the maximum of the energy curve lies at a wave-length of about 0.008 mm., while for carbon at the temperature of the arc it has shifted to the edge of the visible spectrum, and for the sun it is in the yellow.

The emission (E_λ) for any wave-length for a black body is given with very great accuracy by the expression

$$\log E_\lambda = K_1 + \frac{K_2}{T} \qquad (3)$$

Where K_1 and K_2 are constants and T is the absolute temperature. For some other radiating surfaces E_λ has been found to follow quite closely the same law, though with different constants.

339. Radiation Pyrometry.—By this is meant the measurement of high temperatures by observing the variation with temperature of either total emissivity or partial emissivity. In the former case radiation of all wave-lengths from the surface whose temperature is to be measured is allowed to fall on a thermopile of some sort and the resulting deflection of a voltmeter or galvanometer is noted. By observations on a surface at known temperatures the instrument can be calibrated so as to indicate temperatures directly. Instruments of this sort are the Féry or Thwing total radiation pyrometers. If the instrument is *calibrated* by using a perfect radiator, and *used* on another surface, it will indicate, not the true temperature of this surface, but the temperature of a perfect radiator, which would radiate with the same total intensity as this surface. This is called a *black body temperature* of the surface, and will usually be lower (it cannot be higher) than the true temperature.

Optical pyrometry makes use of the partial emissivity. The

method consists in comparing the radiation of a given wave-length (usually red) from the surface whose temperature is desired and from a comparison source, usually a small incandescent lamp. In using the instrument the electric current is measured which is required to heat the comparison lamp so that it disappears when viewed against the hot body. The instrument is calibrated by observations on a black body at known temperatures, and the radiation laws given by the equation on p. 276 may be used to extend the scale beyond the region of possible comparison. In this way measurements have been made up to 3600°C. If used on a surface other than a perfect radiator, it will give a "black body temperature," less than the

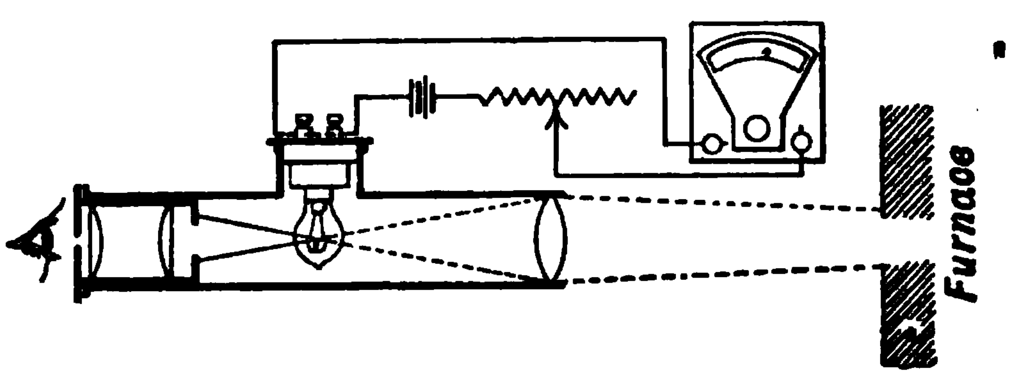


FIG. 213.—Optical pyrometer for high temperature measurements. Ammeter for measuring current in comparison lamp.

true temperature. Examples of this type of instrument are the Holborn, Morse, and Wanner optical pyrometers, the first named being shown in Fig. 213. Radiation pyrometry at present is the only satisfactory method available above about 1750°C.

THE CONSERVATION OF ENERGY

340. The Transformation of Mechanical Energy into Heat.—Since heat is energy and can be produced by the transformation of mechanical energy, it is of great importance to determine just how much mechanical energy is equal to unit quantity of heat. In the *c.g.s.* system, the *mechanical equivalent of heat is the number of ergs equivalent to* (*i.e.*, which will produce) *one calorie*. The symbol for it is J.

The first careful determination of this important quantity was by Joule of Manchester in 1843, before the caloric theory was finally overthrown. Rowland in 1878 carried out one of the most reliable determinations of J which have been made, the method being an improvement of one used by Joule many years

before. A calorimeter (Fig. 214) contains water and one fixed and one movable set of paddles, the latter driven by a shaft through the bottom of the calorimeter, and the former so arranged that the water can not rotate as a mass, but will be violently churned. The paddles are driven at a steady speed by a steam engine, and the calorimeter prevented from rotating by the couple applied through two cords which pass tangentially from a carefully turned rim of radius R, and after passing over frictionless pulleys carry two weights of M grams each. The resisting couple experienced by the paddles in their motion through the water must be equal in magnitude to the couple which the

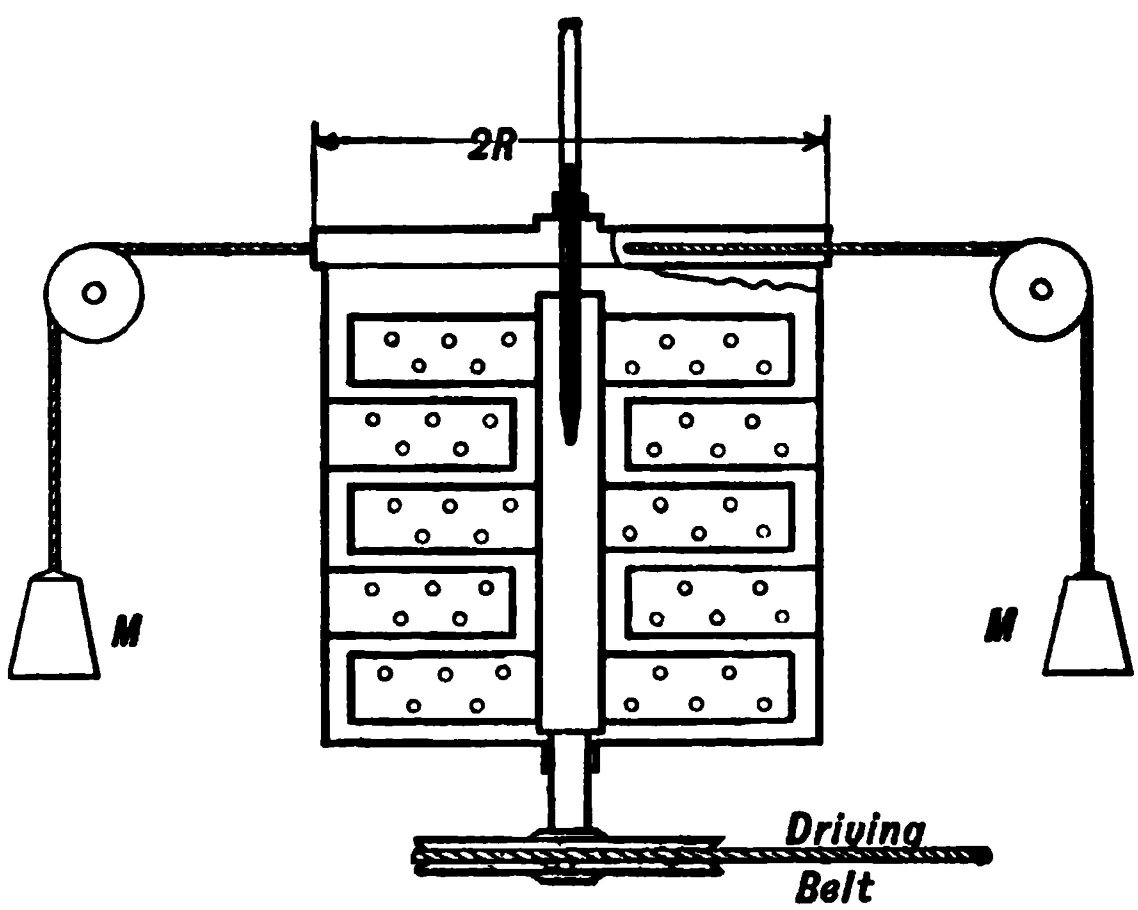


FIG. 214.—Apparatus for measuring the mechanical equivalent of heat.

water exerts tending to rotate the calorimeter. The resisting couple increases as the speed of rotation is increased. If for a given speed the weights M are so adjusted that the calorimeter does not rotate, and if we let L represent the moment of the resisting couple, then

$$L = 2RMg.$$

From this the work done by the paddles can be calculated.

$$\text{Work in one revolution} = 2\pi L = 4\pi RMg,$$

$$\text{Work in } N \text{ revolutions} = 4\pi NRMg.$$

If, after the N revolutions, the temperature of the calorimeter has risen $(t_2 - t_1)°$, and if m and m' are the mass of water contained in,

and the water equivalent of the calorimeter respectively, then $(m+m')(t_2-t_1)$ = the heat added to and retained by the water. To this must be added the heat lost by convection and radiation, which we shall denote by H. There is no correction to be applied to the expression for the work, for only the work done against the frictional forces in the calorimeter is measured. Hence $(4\pi NRMg)$ ergs are equivalent to $(m+m')(t_2-t_1)+H$ calories

or
$$J=\frac{4\pi NRMg}{(m+m')(t_2-t_1)+H} \text{ ergs per calorie}$$

Allowance must also be made for variation in the specific heat of water.

Other methods have been used in which the friction occurred between metal surfaces, the heat being absorbed from them by the water of the calorimeter or by a stream of water flowing past them (see §289). The latter arrangement is identical with one of the standard forms of *absorption dynamometers* used in engineering practice for obtaining the power developed by an engine or motor. Since the mechanical value of electrical work can be very accurately determined (§458), it is possible to determine J indirectly by converting electrical energy into heat, as has been done by Callendar and others. This method is simpler than the direct one. The average of the four best determinations is

$$J=4.187\times 10^7\frac{\text{ergs}}{\text{Cal}_{15}}$$

which is practically Rowland's value and is probably correct to $\frac{1}{10}$ per cent. The numerical value of J depends, of course, on the units, the following being also used

$J=427$ kilogram-meters per large calorie.
$J=778$ foot-pounds per B. T. U.

341. The Law of Conservation of Energy.—We have already become familiar in Mechanics with the transformation of kinetic energy ($\frac{1}{2}mv^2$) into potential energy when work is done against mechanical forces, and we have given reasons for believing that heat is a special form of kinetic and potential energy.

Later we shall have to deal with electric and magnetic forces and work done against them, giving us the idea of electric and magnetic kinetic and potential energy, while chemistry deals with chemical potential energy, though this may ultimately be found to be electrical in nature.

After the growth of the idea that heat is energy, and Joule's early (1843) determination of J, Helmholtz, in 1847, formulated the idea that not only heat and mechanical energy, but all forms of energy are equivalent, and that a given amount of one form cannot be made to disappear without an equal amount appearing in some of the other forms. For example, when the potential energy of a wound clock spring disappears, heat, caused by work done against frictional forces, appears in the clock, while energy of sound waves and kinetic energy of motion of parts of the clock are also produced. Again, the heat energy of steam may be transformed into mechanical energy by a steam engine and given to a dynamo, which does work against electric and magnetic forces, producing some heat but largely electric potential energy, which in turn is changed, by the flow of an electric current, partly into heat in the wire, but largely into mechanical work by a motor, or into light and heat by an electric lamp. This idea of equivalence may be expressed in many ways, such as,

Energy is indestructible.

The total amount of energy in the universe is constant.

The energy required to change a system of bodies from one state (including of course its electric and magnetic condition) to another state is independent of the particular intermediate states through which it passes.

These are all statements of the *Law of Conservation of Energy*, of which the last is perhaps the best, because we cannot deal with the universe, nor can we measure the *total amount* of energy present in any body. The fundamental idea is that all processes, such as the change of the energy of steam into mechanical energy and light above mentioned, consist in drawing a stream of energy from some source and then dividing and diverting that stream into various channels such as heat, mechanical work, light, etc. Common experience shows us that it is always very easy to convert any other form of energy into heat. Whenever a bell is

rung by a battery, or a pump operated by a wind mill, *some* of the energy of the battery or the wind is changed into heat.

Like all the greatest fundamental physical laws, the law of conservation of energy is not capable of direct proof, but is a *hypothesis consistent with all known facts*, which is to be accepted until some phenomena are discovered with which it is inconsistent. It is of the widest possible application and is the chief basis of all physical, astronomical and chemical reasoning, as well as of engineering practice. It leads us to doubt at once all "perpetual motion" devices which purport to obtain mechanical work from nothing.

THERMODYNAMICS

342. First Law of Thermodynamics.—Thermodynamics is the analysis and discussion of the problems of converting heat into other forms of energy, and other forms of energy into heat, and consists in the deduction of consequences from two very general principles, the first one being the law of *conservation of energy* (§341).

Considering a body or a system or group of bodies as distinct from its surroundings, we have already (§262) defined the term "internal energy," for which we shall use the symbol U, as the entire energy which the system contains. As was pointed out earlier, we have no knowledge of the value of U in any case, but we can study the *changes* in U. If the reactions between the system and outside bodies are such as to permit the passage of heat to or from the system, and the doing of work on or by the system, then it follows from the law of conservation of energy. that for any change in the system

$$\left.\begin{matrix}\text{the increase in}\\ \text{internal energy}\end{matrix}\right\} = \text{the heat added} + \left\{\begin{matrix}\text{the work done}\\ \text{on the system.}\end{matrix}\right.$$

This is in fact merely a generalization (applied to a system of bodies instead of to a body) of the statement of §292 that the heat added to a body = the increase in internal energy + external work done by the body. The essential idea is that the energy added must all be accounted for—no part of it is lost.

343. Isothermal Processes.—Any process or change of condition in a system which takes place without change in temperature is an isothermal process. We must distinguish between an

isothermal process and an isothermal curve for a substance. Suppose the substance is in the gaseous state, then we have seen (§281) that at a given absolute temperature T the possible pressures and volumes are given very approximately by $PV = RT$, the isothermal curves being rectangular hyperbolæ. A gas having the pressure and volume determined by this equation, at the given temperature, would, if confined in a cylinder with movable piston, be in *equilibrium*, that is, the piston would not move. All the conditions determined by the equation $PV = RT$ or the corresponding isothermal equation for a substance not a gas, are *equilibrium* conditions, the pressure being the equilibrium pressure corresponding to the given volume and temperature. It is evident that to make a substance *change* its condition (T constant) the confining pressure must be changed from the equilibrum pressure, increased if it is desired to compress the gas, diminished if the gas is to be allowed to expand. If the change in pressure is very slight the change in condition is slow; if the pressure is kept continuously slightly different from the equilibrium pressure given by $PV = RT =$ constant, the gas will pass through a series of conditions, in this case isothermally The volumes of the gas, and the corresponding pressures exerted by the piston upon it, plotted on the PV diagram (Fig. 215) give the dotted curve just above and below the isothermal curve for the same temperature and by making the process *slow* enough the dotted curve representing it may be made to approach as near as we wish to the equilibrium curve. We have seen that the work done upon the gas during isothermal compression is equal to the area under the isothermal curve between the extreme ordinates, and from the law of conservation of energy it follows that, neglecting the *change in internal energy* with volume, which we have seen (§292) is small, the equivalent of the work done upon the gas must be taken away as heat in order to keep the

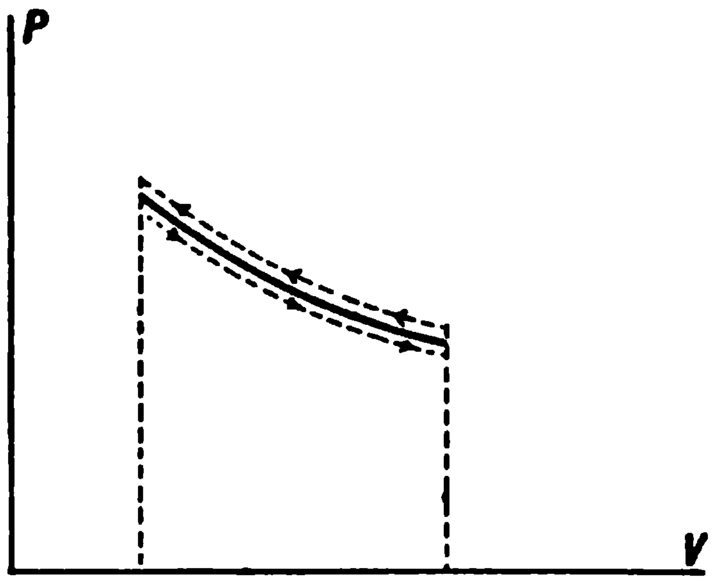


Fig. 215.—Isothermal equilibrium curve, and curves of reversible isothermal process.

temperature constant. In a similar way, during isothermal expansion heat must be added.

344. Adiabatic Processes.—Any process carried out in such a way that no heat is allowed to enter or leave the system during the change, is called an adiabatic process. The same distinction as before exists between a process and a curve, an adiabatic curve determining a series of equilibrium conditions.

Through every point on the PV diagram one adiabatic and one isothermal curve will pass, the adiabatic being everywhere steeper than the isothermal, because, since no heat is added, the temperature of the gas will fall as it expands and does work. Conversely, a substance has its temperature raised by adiabatic compression, the heat equivalent of the work done remaining in the substance, the work done being always represented by the area under the adiabatic curve between the extreme ordinates. The difference between isothermal and adiabatic compression may easily be illustrated by the use of a good bicycle pump, a slow compression being almost isothermal, the heat passing off as it is generated, through the metal walls of the cylinder, while quick compression warms the gas and cylinder considerably, as will be evident to the touch. Since it is impossible ever entirely to eliminate loss of heat by conduction, convection, or radiation, *quick* changes of volume will in general be more nearly adiabatic than slow. The compressions of sound waves are adiabatic for this reason.

The equation of an adiabatic curve of a perfect gas, in the PV diagram, is

$$PV^{\kappa}=\text{constant}$$

κ being the ratio of the two specific heats, $\frac{s_p}{s_v}$. The same equation, having the same meaning, also holds very approximately for *real gases* which closely follow Boyle's law; and even for CO_2 which departs very considerably from Boyle's law, the adiabatic curve is given by the same *form* of equation, though κ is not in this case the ratio $\frac{s_p}{s_v}$.

345. The Equation of an Adiabatic.—Consider 1 gram of a gas, which obeys Boyle's law at least approximately, confined in a cylinder with a movable piston.

Let V = the initial volume of gas.
P = corresponding pressure of the gas.
T = temperature of the gas.

On the PV d agram, Fig. 216, this condition will be represented by the point (a). Let T be the isothermal curve (temperature T) through this point. Let the piston be moved so as to compress the gas an amount $ab = \Delta V$. If no heat is allowed to enter or leave the gas during this compression, then, by definition, the pressure and volume after compression determine a point (d) on the adiabatic through (a). If, on the other hand, the temperature is maintained constant during compression, then the final condition is the point (c). In the first case the work done on the gas is measured by the area $(adef)$, and it is evident that the less ΔV, the more nearly will this area be equal to the area $(abef)$. Hence for small values of ΔV, we can

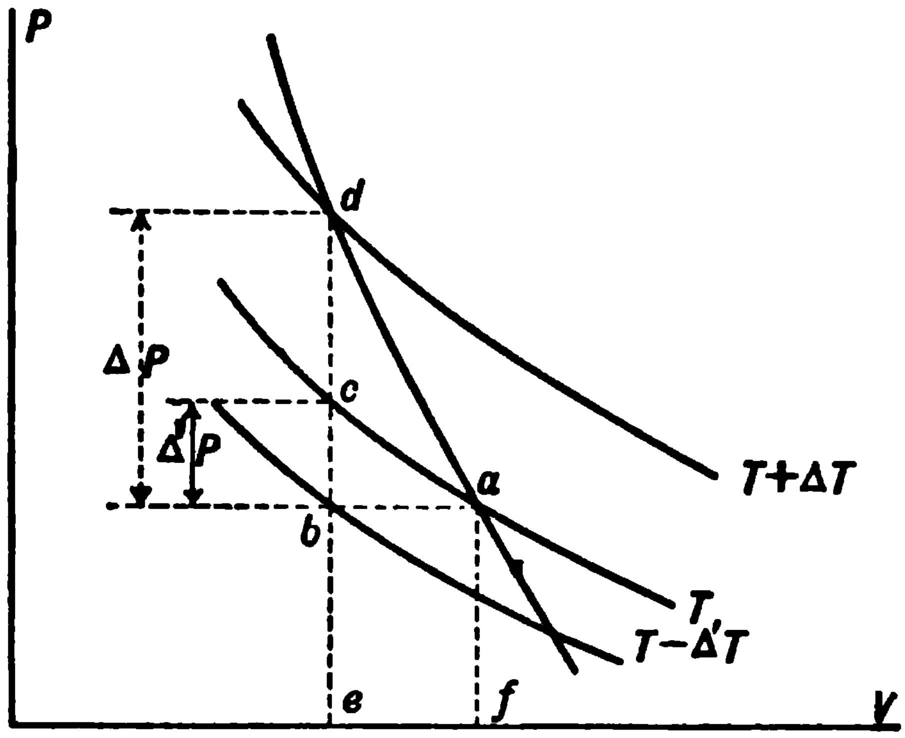


FIG. 216.—A small adiabatic change analysed into the equivalent "volume constant" and "pressure constant" changes.

substitute for the direct compression (a, d) the steps (a, b) and (b, d), and it follows from the definition of an adiabatic that

the heat given out in step (a, b) = the heat added in step (b, d)

or

$$s_P \Delta'T = s_V (\Delta T + \Delta'T) \tag{1}$$

But the change in pressure of a gas confined at constant volume is approximately proportional to the change in temperature; hence, from (1)

$$\frac{s_P}{s_V} = \frac{\Delta T + \Delta'T}{\Delta'T} = \frac{\Delta P}{\Delta'P} = \kappa \tag{2}$$

where κ denotes the ratio $\frac{s_P}{s_V}$

But, since (c) is on the isothermal through (a), and ΔV is negative

$$(P + \Delta'P)(V + \Delta V) = PV,$$

(see §223) or

$$V\Delta'P + P\Delta V = O \tag{3}$$

and, substituting from (2) for $\Delta' P$,

$$\frac{V\Delta P}{\kappa}+P\Delta V=O$$

or, for infinitesimal changes, we have

$$\frac{dP}{P}+\kappa\frac{dV}{V}=O \tag{4}$$

and, integrating,

$$PV^{\kappa}=\text{constant.}$$

346. Adiabatic Elasticity of a Gas.—The modulus of volume elasticity of a gas has been defined in §169 as

$$E=-\frac{dP}{dV}V$$

From equation (4) of §345 we see at once that for an adiabatic change of volume

$$E_{Ad}=-\left(\frac{dP}{dV}\right)_{Ad}V=\kappa P$$

347. Cyclic Operations.—A cyclic operation or cycle, is a process or a series of processes so arranged that the system undergoing these changes is finally brought back to its initial condition. On the PV diagram any closed curve would evidently represent a cycle. Any such cycle may be divided into an expansion and a contraction (Fig. 217), and the area under the curve ABC represents the work done *by* the substances in the expansion ABC, while the area under the curve CDA represents the work done *on* the substance, during the compression CDA. The net work, in this case done *by* the substance, is evidently the area enclosed by the curve $ABCD$. If the cycle were described in the opposite sense, $ADCB$, the same amount of net work would be done *on* the substance. These conclusions are entirely general.

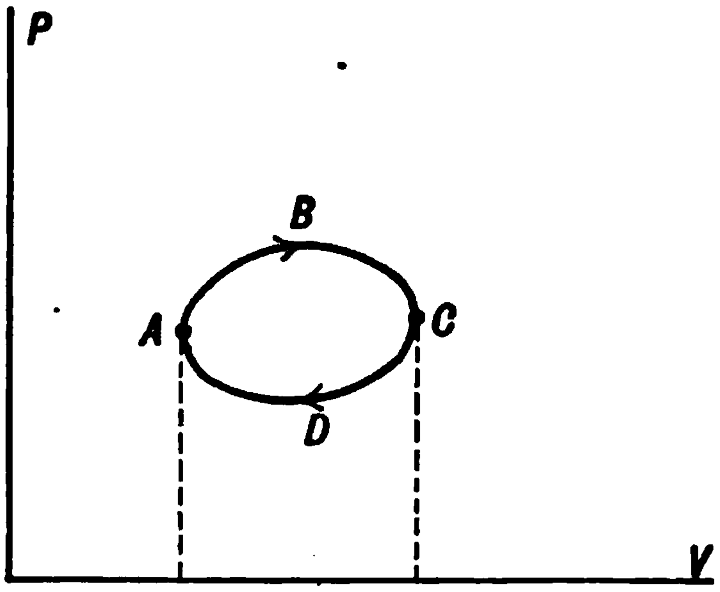


FIG. 217.—Curve representing a cyclic operation.

348. Reversible Processes and Cycles.—Any process is defined as reversible if it can be made to take place in the opposite sense by an infinitesimal change in the conditions, or, what is the same thing, if the curve representing the *process* (§343) lies infinitesimally near an equilibrium curve. For example, to make an isothermal process reversible, the pressure during expansion

must always be infinitesimally near but less than, and during compression, infinitesimally near but greater than, the equilibrium pressure given by $PV = RT$, and the flow of heat must take place under an infinitesimal temperature gradient, that is *to* a body whose temperature is dT lower than that of the gas, or *from* a body whose temperature is dT greater than that of the gas. Under these conditions infinitesimal changes in P and T will cause the process to be described in the opposite direction. A cycle will be reversible if it is entirely made up of reversible processes.

349. The Carnot Cycle.—Carnot's Cycle, Fig. 218, is made up of two isothermal and two adiabatic processes so chosen that the initial and final states are the same. Given a material, called the "working substance," conveniently (though not necessarily) a gas, inclosed in a cylinder with non-conducting walls and piston and a good conducting bottom (Fig. 219), together with a body (1) of very large heat capacity, at temperature T_1, a non-conducting stand (S), and a second body (2) of large heat capacity at temperature T_2, the Carnot cycle may be carried out as follows:

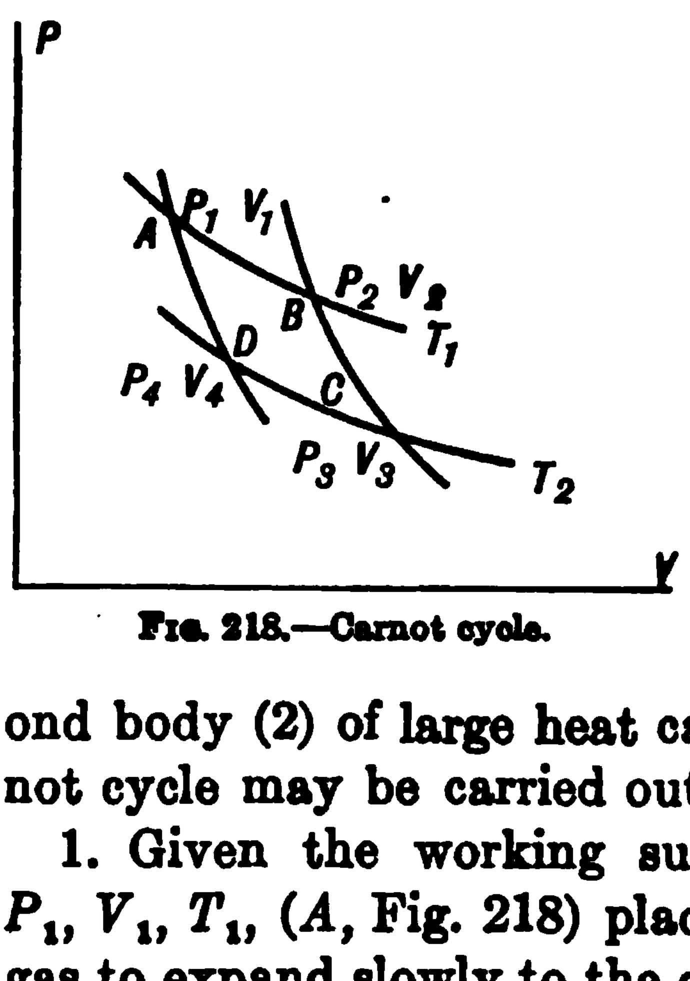


FIG. 218.—Carnot cycle.

1. Given the working substance initially in the condition P_1, V_1, T_1, (A, Fig. 218) place the cylinder on (1) and allow the gas to expand slowly to the condition P_2, V_2, T_1, absorbing heat by conduction from (1) during the process. If done slowly the process will be *isothermal* and *reversible*.

2. Place the cylinder on the insulated stand S, and allow the working substance to expand adiabatically (and reversibly) to the condition P_3, V_3, T_2.

3. Place the cylinder on the refrigerator (2) and compress isothermally to the condition P_4, V_4, T_2, heat being given off during the process to body 2. This will also be reversible if the compression is slow.

4. Place the cylinder again on the insulating stand and compress adiabatically to the initial condition.

According to §347 the net work (W) done by the gas when the cycle is described in this sense is represented by the area $ABCD$.
Let H_1 = heat taken in at temperature T_1 in mechanical units.
H_2 = heat given out at temperature T_2 in mechanical units.
Then according to the first law

$$W = H_1 - H_2$$

If the cycle were carried out in the reverse order, then—

H'_1 = heat given out at temperature T_1
H'_2 = heat taken in at temperature T_2
W' = work done on the gas during one cycle

and again

$$W' = H'_1 - H'_2$$

and $W = W'$

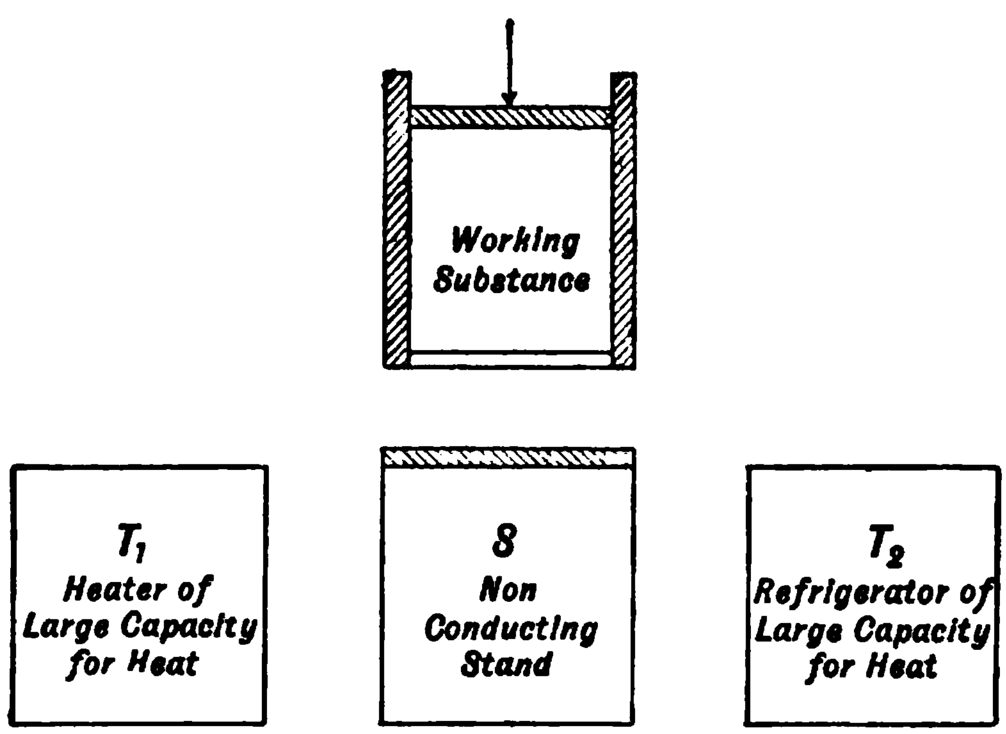


FIG. 219.—Carnot "engine"; a device for using the Carnot cycle.

In order that the cycle may be reversed it is necessary, as we have seen, that the heat flow should take place with infinitesimal temperature gradient, and the pressures always be infinitely near equilibrium pressures. The first condition can be satisfied as near as we wish by making the isothermal transformations slow enough and the second condition by properly altering the force on the piston. The process which we have described, which enables us by means of a reversible Carnot cycle to get mechanical work from heat, is called an *ideal heat engine.* We are not concerned at present with the mechanical construction of such an engine; the essential characteristic is the reversible Carnot cycle, and one ideal engine differs from another only in the material used as working substance, and in the temperatures and pressures

at which it works. An engine working between the temperatures T_1 and T_2 in which the flow of heat to and from the working substance took place under finite temperature gradients, or in which the forces on the piston were not properly adjusted during expansion and compression, or both, would be irreversible.

350. Efficiency of an Engine.—The efficiency of an engine is the ratio of the mechanical work obtained to the heat taken in by the working substance, during one cycle. For the Carnot cycle this is

$$e=\frac{W}{H_1}=\frac{H_1-H_2}{H_1}$$

The efficiency gives the fraction of the heat taken in which is transformed into mechanical work.

351. The Second Law of Thermodynamics.—The second general principle of thermodynamics was first formulated independently by Clausius (1850) and Kelvin (1851) in equivalent but different forms, as follows:

It is impossible for a self-acting machine to convey heat from one body to another at a higher temperature (Clausius).

It is impossible by means of any continuous inanimate agency to derive mechanical work from any portion of matter by cooling it below the lowest temperature of its surroundings (Kelvin).

These are equivalent *axioms* or *assumptions*, which it is impossible to prove directly, but which are to be accepted as a basis of reasoning until some deduction from them is found to contradict fact. No such contradiction has ever been found. The second law recognizes and expresses a certain *natural tendency* of events, for example the tendency of heat to flow *down* a temperature gradient—of a compressed gas to *expand*. Stated in another way it expresses the easily accepted generalization that *natural* processes—that is, processes which take place without assistance or control, are in general *irreversible*, as we have used the term.

352. Carnot's Theorem.—We can now prove an extremely important theorem, which was first stated in 1824 by Carnot as follows:

The efficiency of all reversible engines taking in and giving out heat at the same two temperatures, is the same, and no irreversible engine working between the same two temperatures can have a greater efficiency than this.

Carnot's proof of this theorem was incorrect, being based on the caloric theory of heat. As given by Clausius and Kelvin it is a necessary consequence of the second law. First consider any two reversible ideal engines, E and E' working between the temperatures T_1 and T_2, and let E' run *backward*. Let H_1 and H_2 be, as before, the heat taken in and given out by the forward-running engine, and H'_1 and H'_2 the heat given out and taken in by the engine running *backward*. Also let the engines be so connected mechanically, and of such a size or speed that the work done by the forward-running engine just suffices to operate the backward-running engine. Finally, let us assume for the moment that the efficiency of the forward-running engine is greater than the efficiency of the backward-running one. Then

$$e = \frac{H_1 - H_2}{H_1} > \frac{H'_1 - H'_2}{H'_1} = e' \tag{1}$$

from the inequality of efficiencies, and

$$W = H_1 - H_2 = H'_1 - H'_2 = W' \tag{2}$$

from the equality of the work done by and on the engines, respectively. Hence from (1) and (2)

$$\frac{1}{H_1} > \frac{1}{H'_1}$$

or $$H_1 < H'_1$$

and from (2) $$H_2 < H'_2$$

Hence, the net result is that an amount of heat equal to

$$H'_2 - H_2 = H'_1 - H_1$$

is transferred from the body at the lower temperature T_2 to the body at the higher temperature T_1, without the *necessity of doing any work*. This violates the Clausius statement of the second law—hence we conclude that e cannot be greater than e'. If we run engine E' forward and E backward, we can prove by exactly similar reasoning that e' cannot be greater than e, hence it follows that $e = e'$, which proves the first part of the theorem.

If engine E is an *irreversible* engine, then we can prove exactly as above that e_{ir} cannot be greater than e', but since E cannot be reversed, we *cannot* prove that e' cannot be greater than e_{ir}.

Hence all we can say is that e_{ir} is equal to or less e_{rev},

$$e_{ir} \leqq e_{rev}$$

which proves the second part of the theorem.

353. Thermodynamic Scale of Temperature.—Since the efficiency of a reversible engine is independent of the working substance and the pressures used, it follows that the efficiency can depend only on the two temperatures between which the engine works. If $\frac{H_1-H_2}{H_1}$ depends only on the temperatures T_1 and T_2, then $\frac{H_2}{H_1}$ also depends only on the temperatures. This fact led Lord Kelvin to suggest a new scale of temperature, which, since it depends on Carnot's theorem and is independent of the properties of any particular substance, is called the *absolute thermodynamic scale of temperature*. According to this scale, *any two temperatures are to each other as the heat taken in and given out by a reversible engine describing a Carnot cycle between these two temperatures*. That is, if we call θ_1 and θ_2 the thermodynamic measure of two temperatures, $\frac{\theta_2}{\theta_1}=\frac{H_2}{H_1}$. We still have to determine the size of the degree, which is done, as in the case of the hydrogen scale by assuming 100° between the freezing- and the boiling-point of water, Fig. 220. This amounts to dividing the area (Fig. 220) between the 0° and 100° isothermals and any two adiabatic curves as shown, into one hundred equal parts.

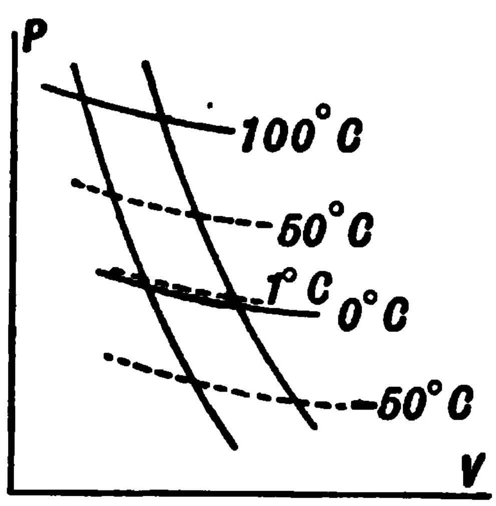


FIG. 220.—Thermodynamic temperature scale; difference in temperature proportional to work done (area).

We have then $$\theta_{100}-\theta_0=100$$

and by definition $$\frac{\theta_{100}}{\theta_0}=\frac{H_{100}}{H_0}$$

hence $$\frac{100+\theta_0}{\theta_0}=\frac{H_{100}}{H_0}$$

and $$\theta_0=\frac{H_0}{\frac{H_{100}-H_0}{100}}$$

that is, the thermodynamic temperature of 0° centigrade is obtained by dividing the heat given out by the reversible engine at 0°C. by $\frac{1}{100}$ of the work done in the cycle from 100°C. to 0°C. Similarly for any other temperature θ we have

$$\theta=\theta_0\frac{H}{H_0}=\frac{H}{\frac{H_{100}-H_0}{100}} \qquad \text{(from above)}$$

which can be interpreted in the same manner. *Absolute thermodynamic zero* is the temperature at which *no* heat is given out by a reversible engine working with this as its lower limit. From the definition we see at once that differences of thermodynamic temperatures are proportional to the work done by an ideal engine working between these limits, and hence to area on the PV diagram (Fig. 220). We may also obtain an expression for efficiency in terms of thermodynamic temperature, namely

$$e=\frac{H_1-H_2}{H_2}=\frac{\theta_1-\theta_2}{\theta_1}$$

and since experiment shows that θ and T are practically the same, we have, approximately

$$e=\frac{T_1-T_2}{T_1}$$

If there were a reversible ideal engine actually available the method of determining thermodynamic temperatures would be first to work the engine between boiling water and melting ice and determine the amount of work it could do, then to work it between zero and a source of available temperature, such as a large tank of water, and adjust the temperature of the tank until the work done was $\frac{1}{100}$ of the amount done from 100° to 0°C. The tank would then be at +1°C. on the *thermodynamic scale.* A similar process would determine other temperatures.

354. Comparison of Thermodynamic and Hydrogen Scale.—The *thermodynamic scale* is entirely distinct from the hydrogen scale, and if it is to be adopted as the standard we must have either a practicable way of measuring in terms of it, or a way of comparing other scales with it.

It can be proved theoretically that the temperature indicated by a gas thermometer operating with a perfect gas would agree exactly with the thermodynamic temperature as defined above, using the perfect gas as the

working substance. Unfortunately there is no perfect gas available for use in a thermometer; but as we have already pointed out, the properties of real gases approach those of a perfect gas as their densities approach zero. Accordingly, if a given gas, for example hydrogen, is used in a thermometer at several densities, and the corresponding temperature scales are compared, the scale obtained by extrapolating from these to a condition of zero density, will agree with the thermodynamic scale. Moreover, real gases differ from perfect gases in several important ways, namely:

$$\left.\begin{matrix}\text{perfect}\\ \text{real}\end{matrix}\right\}\text{ gases obey the law } PV=RT\left\{\begin{matrix}\text{exactly}\\ \text{approximately}\end{matrix}\right.$$

$$\text{the internal work of free expansion is}\left\{\begin{matrix}\text{zero}\\ \text{not zero}\end{matrix}\right\}\text{for}\left\{\begin{matrix}\text{perfect}\\ \text{real}\end{matrix}\right\}\text{gases}$$

$$\begin{matrix}\text{and the specific}\\ \text{heat at constant pressure}\end{matrix}\text{ is}\left\{\begin{matrix}\text{constant}\\ \text{not constant}\end{matrix}\right\}\text{for}\left\{\begin{matrix}\text{perfect}\\ \text{real}\end{matrix}\right\}\text{gases}$$

Hence, by measuring the pressure and volume of a real gas at various constant temperatures, by performing the porous plug experiment (§296) with it, and by measuring its specific heat under various conditions, it is possible to determine its "degree of imperfection," so to speak, and thence the relation between its constant volume temperature scale and the thermodynamic scale. There is still lacking much information concerning the properties of real gases, especially concerning the internal work of free expansion, due to molecular forces. Nevertheless the reduction to the thermodynamic scale is known with considerable accuracy for both hydrogen and nitrogen, as given in Table 17.

TABLE 17

CORRECTIONS FOR CONSTANT VOLUME THERMOMETER SCALES

$P_0=1000$ MM. HG.

Temperature (Centigrade).	Nitrogen.	Hydrogen.
− 240		+0.18
− 200	+0.62	+0.06
− 150	+0.26	+0.033
− 100	+0.10	+0.010
− 50	+0.03	+0.005
+ 10	−0.002	−0.000
+ 40	−0.006	−0.001
+ 70	−0.004	−0.001
+ 200	+0.04	+0.004
+ 450	+0.19	+0.02
+1000	+0.70	+0.07
+1200	+1.00	

It is evident that for moderate temperatures and approximate work the thermodynamic and hydrogen scales may be considered identical.

355. Entropy.—Returning now to the Carnot cycle we see that as a result of the definition of temperature, we have

$$\frac{H_2}{H_1}=\frac{\theta_2}{\theta_1}\quad\text{or}\quad\frac{H_2}{\theta_2}=\frac{H_1}{\theta_1}$$

That is to say, the ratio of the heat taken in (or given out) to the temperature at which it is taken in (or given out) is the same for all isothermal changes between any two adiabatics. This fact suggested to Clausius that the quantity $\frac{H}{\theta}$ is the change in a certain property of the working substance, a property which remains constant during any (reversible) adiabatic process but changes when the substance passes from one adiabatic to another. This property Clausius named "entropy," and it is exceedingly important.

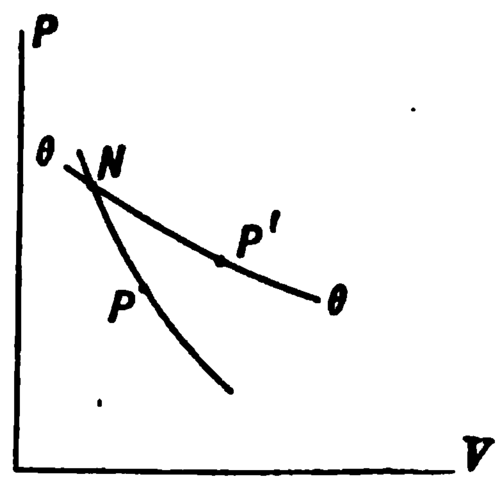


FIG. 221.—Arbitrary zero of entropy, P; Entropy of P' determined by adiabatic-isothermal change from P.

In order to obtain a definite numerical measure for the entropy of a body in every physical condition, we must select some condition, represented by a point on the PV diagram, as an arbitrary zero of entropy, just as we select the sea level as the zero from which to measure heights and depths. Suppose P (Fig. 221) is the adopted zero, then the entropy of any other state P' is obtained by measuring the heat taken in (or given out) in passing from P to P' by a *reversible* path. The simplest path is by the adiabatic PN and the isothermal θ. If H is the heat taken in in passing from N to P', then the entropy of P' with respect to P, which we shall represent by $S_P^{P'}$, would be equal to $\frac{H}{\theta}$. If P' were reached by another reversible path involving portions of several adiabatics and isothermals, and quantities of heat H_1, H_2, $H_3 \cdots$ were taken in (or given out) at the temperatures θ_1, θ_2, θ_3, $\cdots$ then

$$S_P^{P'} = \frac{H_1}{\theta_1} + \frac{H_2}{\theta_2} + \frac{H_3}{\theta_3} = \sum_P^{P'} \frac{H}{\theta}.$$

If any of the quantities of heat H_1, $H_2 \cdots$ were given out by the body they should be taken with the minus sign in the summation. It is evident that, defined in this way, every state has a definite entropy.

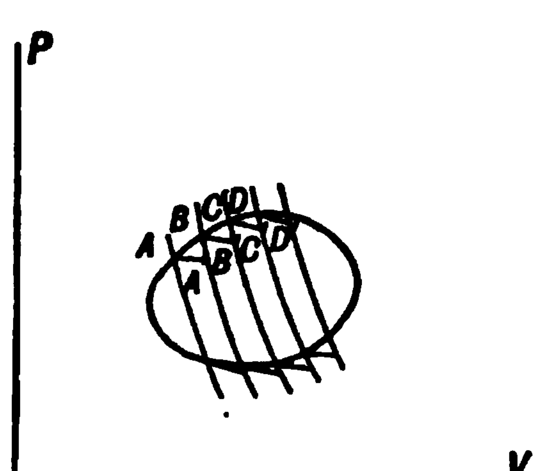


FIG. 222.—Analysis of any cycle into elementary Carnot cycles.

356. Entropy and Reversible Cycles in General.—We have seen that in passing around a Carnot cycle the entropy of the working substance was not changed. This result may be extended to include any reversible cycle, represented by the closed curve in Fig. 222. By drawing a series of adiabatics across this and connecting these around the edge by a series of isothermal steps as shown, we see that the given cycle may be broken up into a series of Carnot cycles the sum of whose areas will approach the area of the given cycle as a limit, as their number is increased. Furthermore, the heat taken in along the isothermal steps AA', BB', CC',

DD', etc., is equal in the limit to the heat taken in moving along the curve A—D, for the difference would be represented by the sum of the triangular areas, which is zero in the limit. Hence, since for each elementary Carnot cycle we have—

$$\frac{H_1}{\theta_1}=\frac{H_2}{\theta_2}$$

we have for the whole cycle $\displaystyle\sum\frac{H_1}{\theta_1}-\sum\frac{H_2}{\theta_2}=0$

or, if we give the negative sign to heat H_2 *leaving* the system, this becomes,

$$\sum\frac{H}{\theta}=0$$

$$\int\frac{dH}{\theta}=0$$

when the number of elementary cycles has become infinite. This shows us that all reversible paths between two conditions involve the same change in entropy, or that $\dfrac{dH}{\theta}$ is a *perfect differential.*

357. Increase in Entropy.—If an amount of heat H flows from one body at a temperature θ_1 to another at a lower temperature θ_2, the entropy of the hot body is decreased by an amount $\dfrac{H}{\theta_1}$ and that of the cooler body is increased by $\dfrac{H}{\theta_2}\cdot$ Evidently in all cases of conduction, $dS=H\left(\dfrac{1}{\theta_2}-\dfrac{1}{\theta_1}\right)$, is *positive*, or the entropy of the two bodies is increased.

It can also be proved that other "natural" processes such as free or unbalanced expansion, diffusion, the falling of bodies in obedience to gravitation, and the production of heat from mechanical energy by friction, all involve an increase in entropy. These processes are also all irreversible, and they all tend to a more uniform condition as regards temperature, pressure and the velocities of bodies and of molecules. Hence it is a reasonable extension of our ideas to say that *all* natural processes are irreversible and lead to an increase in entropy, and to associate the increase of entropy with increase in the *uniformity of physical conditions*. All natural changes seem to be tending to a condition of *maximum uniformity*. This additional hypothesis, that natural processes always lead to an increase in entropy, is the basis for the discussion of problems of chemical and physical equilibrium, such as the equilibrium of a liquid with its vapor, of a solid with its liquid, or of different chemical compounds with each other.

We have seen that for any irreversible cycle during which heat is taken in and given out at temperatures θ_1 and θ_2,

$$\frac{H'_1-H'_2}{H'_1}\lesseqgtr\frac{\theta_1-\theta_2}{\theta_1}$$

from which

$$1-\frac{H'_2}{H'_1}\lesseqgtr 1-\frac{\theta_2}{\theta_1}$$

or $$\frac{H'_2}{\theta_2} \geqq \frac{H_1}{\theta_1} \quad \text{or} \quad \int \frac{dH}{\theta} \leqq 0$$

So that in this general case all we can prove from the second law is that there *may* be an increase in entropy; and by an extension of this reasoning it may be proved that *no* irreversible process can lead to a *decrease* in entropy.

358. Reciprocating Steam Engines.—The ordinary reciprocating steam engine, one type of which is shown in Fig. 223, is the most common machine used to convert heat energy into mechanical work. In these engines water is the *working substance*

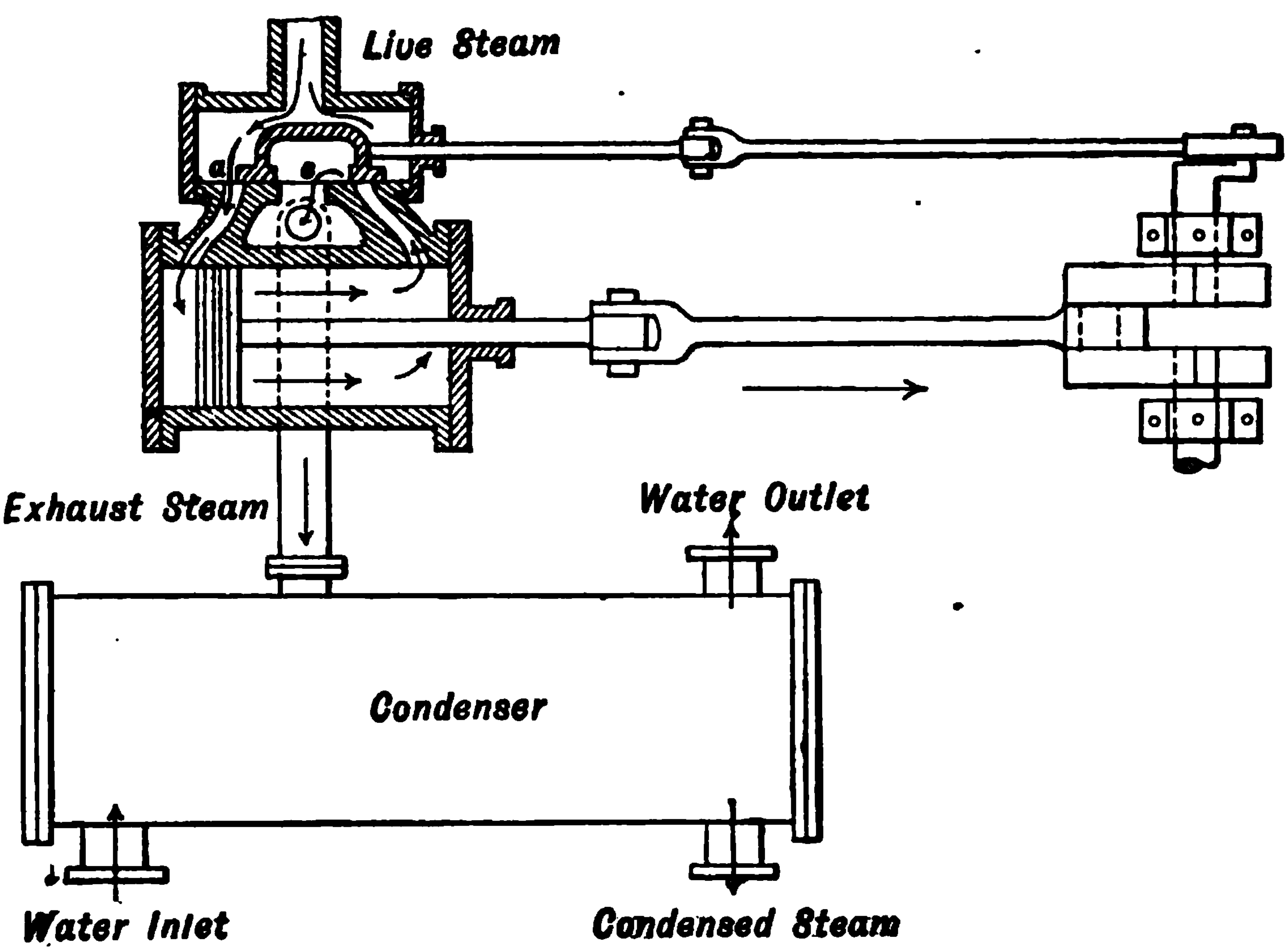


FIG. 223.—Reciprocating steam engine and condenser in which steam is condensed on water-cooled pipes.

(Compare Art. 349), the boiler is the source of heat at the higher temperature, and the cooling water of the *condenser* is the cooler body into which heat is discharged. Instead of moving the cylinder from one to the other as was before suggested, it is obviously easier to conduct the working substance from point to point. On account of mechanical difficulties no attempt is made to realize completely the Carnot cycle (Art. 349), but the actual cycle through which the working substance passes is of the form shown in Fig. 224. The operations are as follows:

(1) Water is vaporized in the boiler at the temperature T_1, absorbing an amount of heat L_1 per unit mass (heat of vaporization).

(2) Steam passes at constant pressure P_1 from the boiler through the valve *a* (Fig. 223), into the cylinder as the piston begins its motion to the right. Thus the isothermal expansion at pressure P_1 due to vaporization is represented by the line ($A,B,$), and this expansion does an amount of work represented by A,B,G,F.

(3) The valve *a* closes, and the saturated steam expands from B to D. This expansion should be as nearly as possible adiabatic. The work done is represented by the area $BDHG$. At C the valve *a* opens to the exhaust *e*, the steam begins to escape to the condenser, and the pressure falls quickly from C to D.

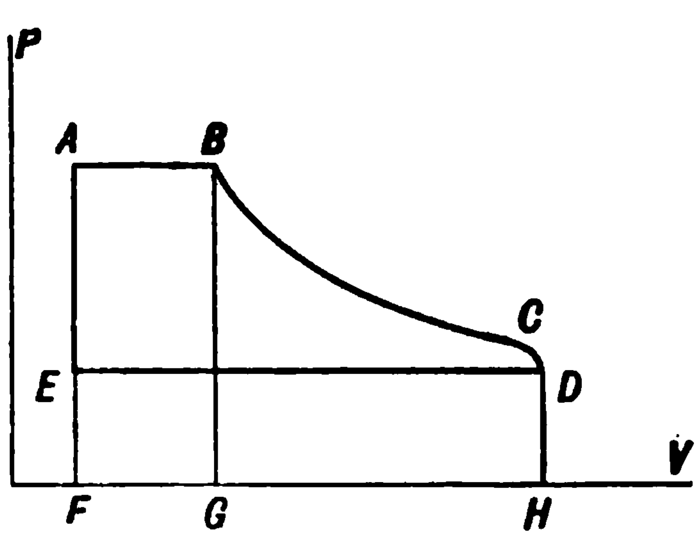


FIG. 224.—Ideal (Rankine) cycle for a reciprocating steam engine.

(4) The piston reverses its motion at D, and the motion to the left is opposed by the constant pressure P_2, since during this time there is isothermal condensation of the steam in the condenser, at temperature T_2. The temperature T_2 is fixed by the cooling water which is available for the condenser. With a non-condensing engine T_2 is necessarily about 373° absolute (100°C.). An amount of heat L_2 per unit mass is given up to the cooling water during the process of condensation, and work to the amount $DEFH$ is done *on* the steam.

(5) The condensed steam is heated at constant volume (E,A) and admitted to the boiler at A, thus completing the cycle. This requires an additional amount of heat H.

It is possible to arrange a mechanism so that the engine as it runs will automatically draw a curve whose ordinates are proportional to the pressure of steam in the cylinder, and whose abscissæ are proportional to the corresponding volume occupied by the steam in the cylinder. This curve is very similar to the one of Fig. 224, and is called an *indicator* diagram.

EFFICIENCY OF ENGINES

359. The work, W, done per pound of steam is evidently represented by the area $ABDE$, while the total heat taken in is L_1+H. The ratio $\frac{W}{(L_1+H)J}$, where J is the mechanical equivalent of heat, is called the *thermal efficiency* of the engine. The thermal efficiency measures the perfection of the thermal processes which the engine uses. There is, of course, energy lost (converted into heat) by friction among the moving parts, so that the actual work, W', which the engine could do in running some machine, is always less than W. The ratio $\frac{W'}{W}$ is called the *mechanical efficiency* of the engine, and its value is a measure of the mechanical perfection of the engine. The product of the two efficiencies, namely, the ratio $\frac{W'}{(L_1+H)J}$, evidently measures the efficiency of the engine in the conversion of heat into *usable* mechanical work, and this will be less than its thermal efficiency.

It is interesting to compare the thermal efficiency $\frac{W}{(L_1+H)J}$ with the efficiency of an ideal Carnot engine working between the same temperatures T_1 and T_2.

Since, according to §354 the constant volume hydrogen scale and absolute thermodynamic scale of temperature are practically identical, the expression for the efficiency of an ideal engine becomes $e=\frac{T_1-T_2}{T_1}$, and this is the maximum efficiency which any real engine could possibly be expected to approach if it works with a boiler temperature T_1 and a condenser temperature T_2. For example, with a boiler at 177°C. and a condenser at 77°C., $e=\frac{100}{450}=22$ per cent., that is to say, the ideal engine could convert less than *one-quarter* of the heat used into mechanical work. Table 18 gives the actual thermal efficiency and the corresponding ideal efficiency for the best engines of several types.

Besides engine efficiencies, the efficiency of boilers, namely, the ratio $\frac{\text{heat given to water}}{\text{heat obtained from fuel}}$ (in a given time) is of course

of equal importance in the problem of obtaining mechanical work from fuel. The average efficiency of boilers is 60 per cent., the maximum 80 per cent., so that, combining the best boiler with the best engine, the maximum efficiency actually attained is about 21 per cent.

From §298 the heat of combustion of soft coal is 2.9×10^{11} ergs per gram, or 12,500 B. T. U. per pound, while (§57) one horse-power for one hour equals 2.68×10^{13} ergs, or 1.98×10^{6} foot lbs. Since 1 B. T. U. equals 778 foot lbs., it follows that the combustion of 1 lb. of coal liberates energy sufficient to provide 4.8 H. P. for one hour, whereas the best boiler-engine combination so far built obtains 0.82 H. P. hour per lb. of coal.

TABLE 18

EFFICIENCY OF STEAM ENGINES

	Temperature t_1	Temperature t_2	Efficiency	Efficiency of Carnot cycle
			Per cent.	Per cent.
Willan's engine (non-condensing)	164°	101.5°	10.4	14.5
Levitt pumping engine (compound).	181.6°	37.7°	19	31.7
Levitt pumping engine (triple expansion).	191.9°	46.7°	20.8	31.8
Nordberg engine (quadruple expansion).	206.35°	43.1°	25.5	34.0

360. The Defects of Real Engines.—In the discussion of §359 we have neglected several points of importance. For example, the expansion BC can never be strictly adiabatic because the cylinder and piston must be of conducting material. This leads to the condensation of steam in the cylinder. If it is attempted to raise the temperature T_1 so as to increase the efficiency, the cylinder condensation is increased. By using several cylinders (compound, triple and quadruple), allowing part of the expansion to occur in each, the temperature changes in each cylinder, and hence the condensation losses are reduced and it is possible to use higher initial temperatures. Further reduction of condensation loss, and increase of the initial temperature without increase in the initial pressure, is accomplished by *superheating* the steam, by passing it, at constant pressure, through coils of

pipe in the hot flue gases, as shown in Fig. 225. It is then no longer saturated when it enters the cylinder, and the cycle would be represented by different lines on the *PV* diagram.

Saturated Steam *Superheated Steam*

FIG. 225.—Boiler and superheater.

361. Steam Turbines.—The turbine is another type of machine of more recent development, for obtaining mechanical work from the heat energy of steam, the essential features being a rotating shaft with properly arranged blades and fixed nozzels or blades for directing the flow of the steam, which is initially at a high pressure. Turbines may be divided into two general classes. In the first class, called the "velocity" type, steam is allowed to expand at once to the final pressure, in a properly shaped nozzle, so that the jets acquire a high velocity. These jets impinge on the movable blades and cause them to rotate, much as the jets of water impinge on the blades in certain types of water wheels (Fig. 104, §204). By using several sets of movable blades, with fixed passages between for reversing the direction of the steam jet, the drop in velocity is rendered more gradual, and the speed of the turbine shaft need not be so great. In the second class of turbines, called the "pressure" type, shown in Fig. 226, the steam expands gradually through a great many sets of movable and fixed blades, exerting a pressure on each set which causes the movable blades to rotate. Steam turbines have certain mechanical advantages over reciprocating engines, namely, uniform and

high angular velocity, freedom from vibration (hence their desirability for use in steamships), and economy of space. Turbines are slightly more efficient than reciprocating engines for low working pressures, but slightly less efficient at high

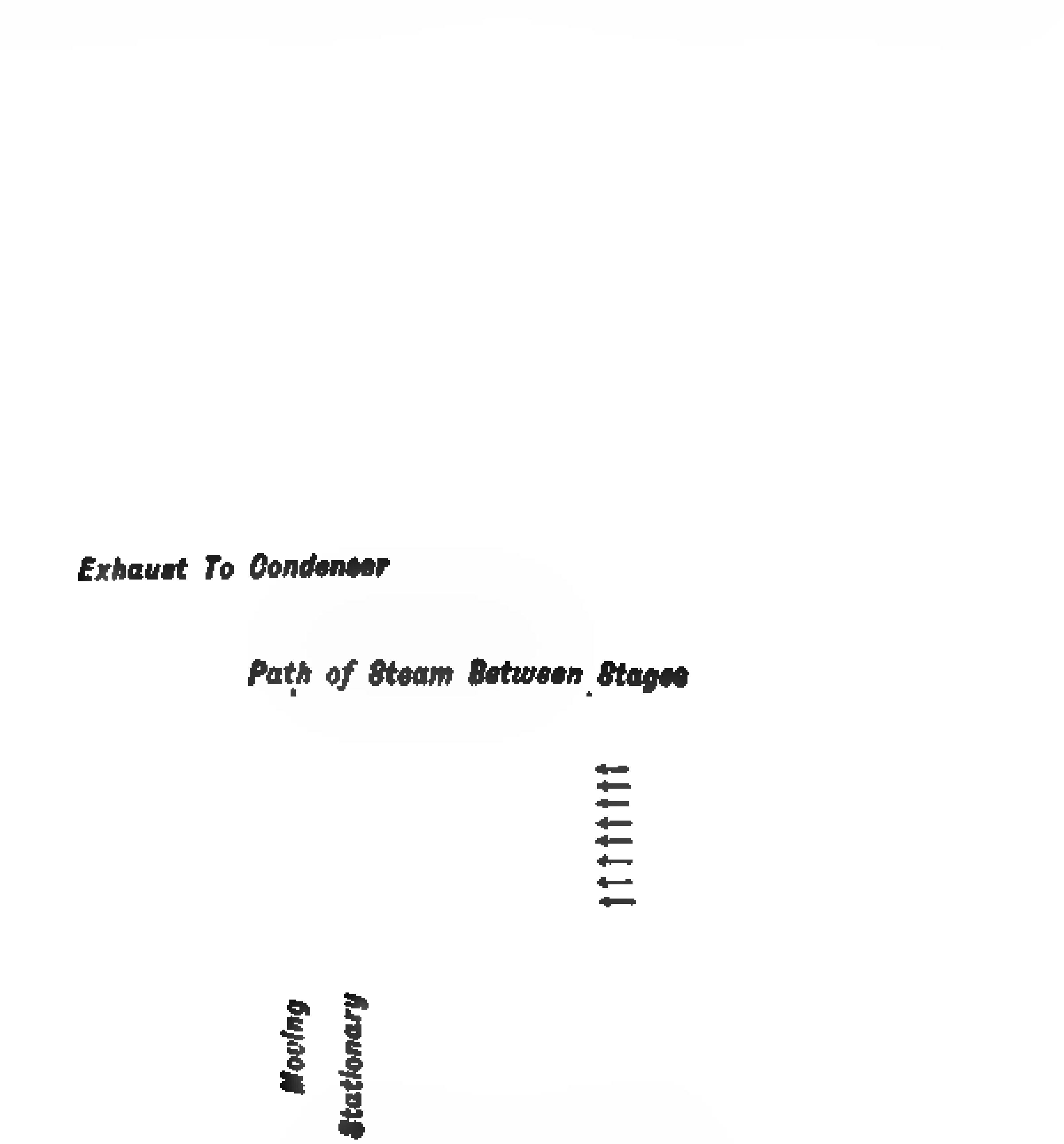


FIG. 236.—Steam turbine, pressure type; general arrangement and detail showing flow of steam past blades.

pressures. Hence it has been found advantageous to combine the two, delivering high pressure steam to a reciprocating engine and allowing the partially expanded steam to pass from it to a low pressure turbine.

362. Internal Combustion Engines.—In these engines the function of the boiler and expanding cylinder are combined, the

combustion taking place in the cylinder itself. The way in which this is carried out in the "four cycle" type of engine can best be understood by describing the various stages shown in Fig. 227. In (1) the inlet valve is open during the entire stroke to the right, admitting a cylinder full of a proper explosive mixture of a com bustible (coal gas, gasoline or alcohol vapor), and air. In (2)

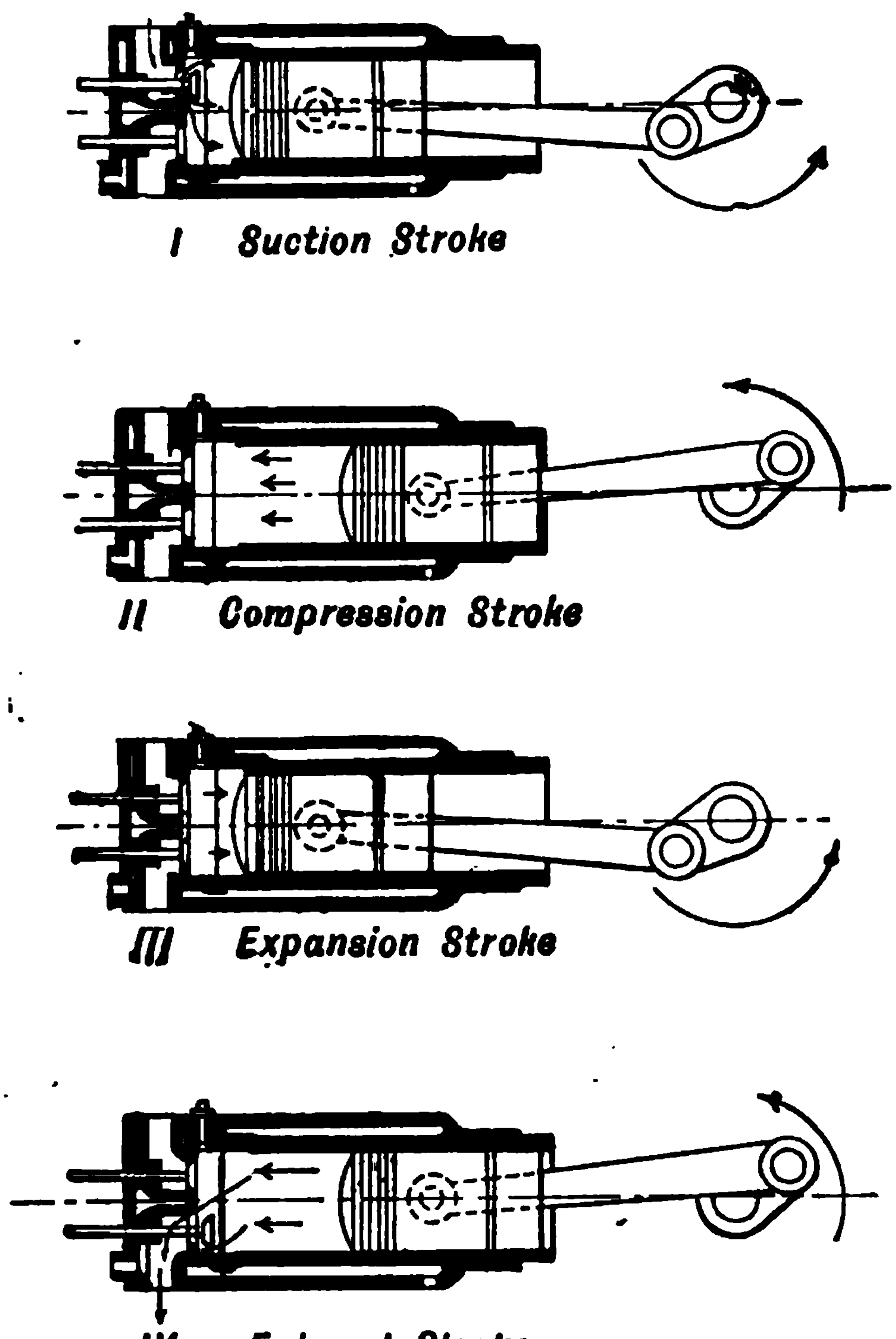


FIG. 227.—Four-cycle internal combustion engine, showing the four stages of one cycle

the valve is closed and the return stroke is taking place; this compresses the mixture into the clearance space at the end of the cylinder, which is called the explosion chamber. At the end of the compression stroke the mixture is exploded, usually by an electric spark. The high pressure resulting from the combustion acts upon the piston during the stroke (3) to the right, while at the end of this stroke the exhaust valve *e* opens and during (4)

the products of combustion are expelled preparatory to beginning over again as in (1). Engines using the series of operations just described are called "four-cycle" engines, because four strokes are necessary to complete the series. There are other types of engines—notably the "two-cycle," requiring only two strokes to complete the series of operations, and the Diesel, in which air alone is compressed and the fuel is injected into it, the result being quiet combustion, instead of an explosion as in the four-cycle type. The thermal efficiency of the best four-cycle engines is about 30 per cent., of the Diesel type about 40 per cent. Aside from high efficiency, internal combustion engines have the further advantages of compactness, ease of handling, and quickness of starting. The formation of the proper mixture of fuel and air is the most troublesome operation in the running of an internal combustion engine, this being usually accomplished in separate attachments called carburetors. The largest engines of the internal combustion type so far built are of 4000 H. P.

References

POINCARÉ, H. *The New Physics.*

A very readable and authoritative book discussing recent developments. Chapters on the various states of matter, and on thermodynamics and the conservation of energy.

TYNDALL. *Heat as a Mode of Motion.*

Popular lectures illustrated by experiments. A beautifully clear and simple presentation.

ENCYCLOPEDIA BRITANICA, 11*th Edition.*

Various articles bearing on the material of the previous section, especially articles "Heat" and "Calorimetry," by Callendar.

GRIFFITHS. *Thermal Measurement of Energy.*

A clear, popular account of the development, from Rumford's time to 1900, of the idea that heat is a form of energy, and of experiments based on this idea.

EDSER. *Heat.*

A text-book covering about the same ground as the previous section, but in greater detail.

POYNTING AND THOMSON. *Heat.*

A text somewhat more advanced than Edser, especially as regards the mathematical treatment.

PRESTON. *The Theory of Heat.*

A valuable reference book, complete and readable. A judicious combination of theoretical and experimental treatment.

DAVIS. *Elementary Meteorology.* An excellent text.

DARLING. *Heat for Engineers.*

A treatment, largely from the experimental side, of various portions of the subject of especial interest to engineers, for example temperature measurement, expansion, combustion, conduction, refrigeration.

GOODENOUGH. *Principles of Thermodynamics.*

The principles of thermodynamics are developed in this book with especial reference to their application to steam, gas, and other heat engines and refrigerating machines.

TRAVERS. *Study of Gases.*

A good chapter on liquefying processes.

BURGESS–LE CHATELIER. *High Temperature Measurements, 3rd. Edition.* The best book in this field.

DAY AND SOSSMAN. *High Temperature Gas Thermometry.* Published by Carnegie Institution.

An account of the most recent methods and results.

"Scientific Memoirs," edited by Ames, is a collection of reprints, or translations, of various important scientific papers, each paper being preceded by a short biographical sketch of the author. The following are of interest:

AMES. *The Free Expansion of Gases.*

Contains papers of Gay-Lussac (1807), Joule (1845), and Kelvin and Joule (1853–1862), all bearing upon the changes in temperature produced by changes in density of gases. The papers of Kelvin and Joule contain the account of the famous "porous plug" experiment.

MAGIE. *The Second Law of Thermodynamics.*

Contains the papers of Carnot (1824), Clausius (1850), and Ke vin (1851), in which the second law of thermodynamics was developed.

The following papers are also of special importance:

BRADLEY. *Physical Review*, 9, 1904, pp. 330–387.

Interesting results of efficiency tests of a liquid air plant.

WAIDNER AND DICKENSON. Bull. Bur. Standards 3, 1907, p. 663, contains descriptions of thermometers and methods of use between 0° and 100°C.

BARNES. Trans. Intern. Electr. Congress, St. Louis, I, 9105, p 53. A good summary of the determination of the *mechanical equivalent of heat* by electrical methods.

WAIDNER AND BURGESS. Bull. Bur. Standards, I, p. 189.

A full description of the methods and instruments of pyrometry.

BOUSFIELD AND BOUSFIELD. Phil. Trans. Royal Soc. of London, 211, 1911, p. 199.

On the specific heat of water and the mechanical equivalent of the calorie. The latest work, and also a summary of previous results.

Problems in Heat

Temperature. 1. Find the value of the following temperatures on the Centigrade scale: The temperature of the human body (98°F.); normal temperature of a living room; a cold day in winter (20°F. below zero). *Ans.* 36.6°C.; 20°C.; –28.9°C.

2. Find the value of the following Centigrade temperatures on the Fahrenheit scale: Absolute zero; melting-point of gold; temperature of sun. *Ans.* –459.4°F.; 1947°F.; 10832°F.

3. At what temperature do the Fahrenheit and Centigrade thermometers read the same? The Fahrenheit twice the Centigrade? *Ans.* –40°; 160°C.

Expansion. 4. A clock which has a pendluum made of brass, keeps correct time at 20°C.; if the temperature falls to 0°C., how many seconds will it gain or lose per day? *Ans.* 16.3 secs. gain.

5. Steel street car rails, having their ends welded together, are laid in concrete so that it is impossible for them to move. Find the stress in the rails at –10°C., assuming that they are laid when the temperature is 20°C. *Ans.* About 7.5×10^8 dynes/cm^2.

6. Compute the change in volume of a block of iron 3 in. × 4 in. × 10 in. if the temperature changes from 44°F. to 116°F. *Ans.* .17 cu. in.

7. A uniform cylinder is filled with hydrogen under atmospheric pressure. The piston stands at a height of 400 cms. at 20°C. If the pressure is kept constant, find the height of the piston at the following temperatures: 100°C.; 300°C.; –80°C.; –180°C. *Ans.* 510 cms.; 783 cms.; 263.4 cms.; 127 cms.

8. If in the preceding problem the same gas is compressed until the piston stands at a height of 200 cms. and the volume is then kept constant, compute the pressure for the temperatures given in problem 7. *Ans.* 193.5; 297.0; 100.; 48.2 cm. Hg.

9. Beginning at 10 atmospheres pressure and 2 liters volume, 10 gr. of air has its pressure so changed that the p, v curve is a 45° straight line. Discuss the temperature changes which occur. (Fig. 187.)

10. Given 10 liters of nitrogen at 30°C. and 120 atmospheres pressure, what would be its volume at 100°C. and 200 atmospheres pressure? *Ans.* 7.4 liters.

11. What would be the relative increase in size of an air bubble in passing from the bottom of a lake 20 m. deep where the temperature is 4°C. to the top where the temperature is 20°C.? *Ans.* $V_2 = V_1 \times 3.1$.

Calorimetry. 12. 100 grams of silver at 100°C. are dropped in 160 grams of water contained in an iron calorimeter weighing 40 grams. Temperature of water initially 15°C. Compute the rise in temperature of water. *Ans.* 2.8°C.

13. 50 grams of a substance at 100°C. are dropped into 100 grams of water at 4°C. If the water is contained in a copper calorimeter, mass 60 grams, and the temperature of the water changes to 10°C., compute the specific heat of the substance. *Ans.* .1403 c. g. s. units.

14. A water heater will heat 50 liters of water per minute from 15°C. to 80°C.; if the efficiency is 25 per cent., how many calories must be generated in the heater to do this? *Ans.* 13×10^6 cals.

15. How many liters of gas will be required per minute in the preceding problem? Density of gas at 0°C. and 760 mm. pressure = .0050. *Ans.* 444 liters.

Mechanical Equivalent of Heat.

16. In drilling a hole in a block of iron, power is supplied at the rate of .8 H. P. for 3 minutes. How much heat is produced? If $\frac{1}{2}$ of this heat goes to warm the iron whose mass is 700 grams, find its change in temperature. *Ans.* 2555 cals; 30°C.

17. How much would the temperature of water be raised by impact after falling 200 ft. under gravity, supposing that all the energy due to its motion was converted into heat. *Ans.* .14°C.

18. Determine the heat produced in stopping a fly-wheel of 112 lbs. mass and 2 ft. in radius, rotating at the rate of one turn per second, assuming the whole mass concentrated in the rim. *Ans.* .364 B. T. U.

19 If electrical energy is 12 cents per 1000 watt hours and gas $1.15 per 1000 cu. ft., what will be the relative cost of gas heating and electric heating? See problem 15 for the density of the gas.
Ans. Cost of elect. = 100 times cost of gas.

Change of State.

20. How much would the air in a room $6 \times 5 \times 3$ meters be warmed by the condensation alone of 1 kg. of steam in the radiator? What would it be if the room were air-tight? *Ans.* 19.4°C.; 27.2°C.

21. With what velocity must a lead bullet at 50°C. strike against an obstacle in order that the heat produced by the arrest of its motion, if all produced within the bullet, might be just sufficient to melt it? *Ans.* 335 m/sec.

22. How much steam at 150°C. must be added to 1 kg. of ice at −10°C. to give nothing but water at 0°C.? *Ans.* 128.5 grams.

23. What is the relative humidity of air at 30°C. if the dew point is found to be 10°C.? *Ans.* 28.7 per cent.

24. How much heat would be required to convert 1 gm. of water-substance from liquid at 0°C. to vapor at 150°C. under 1 atmosphere pressure? *Ans.* 663 cals.

25. Compute the "external" part of the heat of vaporization of water at 100°C. *Ans.* 40.2 cals.

26. Carry a mass of substance across the triple-point diagram as shown on page 253, explaining just what happens at the different points. Do this for both constant pressure and constant temperature following the dotted lines.

27. If it is desired to heat CO_2 at constant volume in a closed tube and have the substance pass through the critical point, what portion of liquid and vapor must there be at 20°C. initially?
Ans. About $4\frac{1}{2}$ parts vapor to 1 of liquid.

28. Some ether is poured into a bottle containing air at atmospheric pressure and the bottle quickly corked; upon shaking the bottle and removing the cork a "pop" is heard. Explain.

Heat Conduction

29. The walls of a certain refrigerator have an area of 15,000 cm.^{2}, and are made of cork 3 cm. thick Find out how much ice may be expected to melt in one day if the outside temperature is 86°F. *Ans.* 21 kg.

30. The top of a steam chest containing steam at atmospheric pressure consists of a slab of stone 61 cm. long, 50 cm. broad and 10 cm. thick. The top being covered with ice, it was found that 48 kilos were melted in 39 minutes. What is the conductivity of the stone?

Ans .0054 c.g.s. units.

31. One end of a copper bar 4 sq. cm. in cross-section and 80 cm. long, is kept in steam under one atmosphere pressure and the other end in contact with melting ice. How many grams of ice will be melted in 10 min.? Neglect loss due to radiation. *Ans.* 34.3 grams.

32. How much anthracite coal must be burned to make up for the loss of heat due to conduction for one day through a glass window 3 mm. thick and having an area of 3 square meters, supposing the air in the room to be at temperature 25°C., and the outside air at -20°C What important point has been neglected? *Ans.* 257 lbs.

Radiation.

33. A black radiator 2 square meters in area, is in a room whose walls are at temperature of 18°C.; if the radiator is at 100°C. at what rate does the room gain heat? The constant s of Stefan's law, $Q = sT^4$, is 5.6×10^{-12} watts per square centimeter.

Ans. 326 cals./sec.

34. If the temperature of a furnace is measured by allowing the heat radiated through a hole 1 square centimeter in area in the walls to warm 100 grams of water placed in front of the hole, what is the temperature if the water rises in temperature by 13°C. in 1 minute? Assume that all the heat radiated from the hole is absorbed by the water and also neglect the heat radiated back into the furnace by the water.

Ans. 1660°C.

35. A blackened copper ball 5 cm.in diameter is heated to 500 °C. and hung by a non-conducting thread in an exhausted vessel. If the absorbing power of the ball and of the walls of the vessel is .98 and the walls are at 0°C., what will be the initial rate of cooling of the ball?

Ans. 4.9°C. per sec.

Thermo-dynamics.

36. How many degrees will dry air at 15°C. rise in temperature if compressed adiabatically to $\frac{1}{5}$ of its volume? *Ans.* 260°C.

37. How much work would be done by air in expanding adiabatically from the point $P = 760$ mm. Hg., $V = 800$ c.c. to the point $P = 400$ mm. Hg.? Solve graphically. *Ans.* 337×10^6 ergs.

38. Plot a curve showing the corrections for a hydrogen thermometer. From your curve find the thermodynamic temperature for the boiling point of oxygen and also the melting point of lead.

Ans. -182.95°C.; 327.013°C.

39. What is the total pressure on the end of a boiler 3 ft. in diameter if the temperature of the water inside is 180°C.? *Ans.* 151,000 lbs.
40. A certain locomotive burns 100 lbs. of soft coal per hour. How much work would the engine do if all this heat were converted into mechanical work? In reality the engine furnishes 20 H. P. What is the efficiency of boiler and engine combined? *Ans.* 1097×10^6 ft lbs.; 3.6 per cent.
41. If an engine working at the rate of 622.4 H. P. keeps a train at constant speed for 10 minutes, how much heat is produced in the rails and bearings, assuming that all the work done is converted into heat? *Ans.* 6.65×10^7 cals.
42. What must be the boiler efficiency in order that a Nordberg quadruple expansion engine should furnish 1 H. P. by burning 1 lb. of soft coal per hour? *Ans.* 88 per cent. for average soft coal.
43. Plot the Carnot cycle with entropy as abscissæ and thermodynamic temperature as ordinate. What does an *area* on this diagram represent? Derive the expression for the efficiency of the cycle.

ELECTRICITY AND MAGNETISM

By Albert P. Carman, Sc. D.

Professor of Physics in the University of Illinois

MAGNETISM

363. Lodestones, Magnets.—Pieces of iron-ore are sometimes found which show a strong and special attraction for particles of iron. When such a piece of iron-ore is dipped into iron filings, the filings cling to it, standing out in tufts, particularly at the edges and at certain points on the piece of iron-ore. A piece of iron-ore which shows this strong and special attraction for iron is called a *lodestone* or *natural magnet.*

By methods which we will study later, a piece of tempered steel, such as a knitting needle or a file, can be magnetized, that is, can be made to acquire the same property as the lodestone for attracting iron. A piece of magnetized steel is often called an *artificial magnet* to distinguish it from the lodestone or natural magnet. It will appear later that there are no essential differences in the properties of "artificial" and "natural" magnets, and we shall accordingly use the term *magnet* for both kinds. Since steel magnets can be had in regular and convenient forms, they are better adapted for showing the properties of magnets and will be used altogether in our study. The two most common shapes given to such magnets are the U-shaped or horse-shoe magnet, and the bar magnet (Fig. 228).

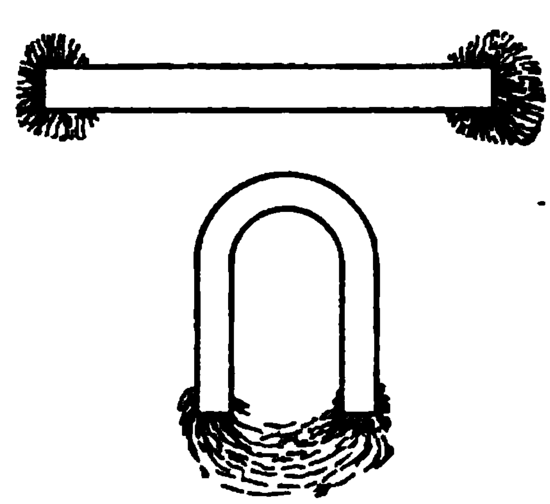


Fig. 228.

Magnetism is a term used for the science of magnets.

364. Magnetic Poles.—When a magnet is dipped into iron filings, it is seen that there are certain points or regions on the magnet of maximum attraction, and other regions where the attraction is zero. A point of maximum attraction is called a

magnetic pole, and a region of no attraction is called a *neutral region.* It is found that a magnet always has at least *two* poles. In a knitting needle as commonly or "normally" magnetized, there are two poles, one near each end of the needle, while the middle of the needle is a neutral region. This distribution is shown by the way the iron filings cling to the needle (Fig. 229). The parts of a magnet showing magnetic attraction are said to have *free polarity*, the total polarity of the middle region being zero (§§369, 375). The straight line joining the two magnetic poles is called the *axis* of the magnet.

FIG. 229.

365. Two Kinds of Poles.—If a normally magnetized knitting needle is suspended so that it rotates freely in a horizontal plane about its middle point, the needle is seen to come to rest in an approximately north-and-south line, with the same pole always pointing to the north, and with the other pole always toward the south. The magnetic pole which points northward is called the *north* (or north-seeking) pole, and the other pole the *south* (or south-seeking) pole. The north pole is commonly designated as the *N* or *positive* (+) pole, and the south pole as the *S* or *negative* pole.

An all important property of magnets is shown by bringing the *N* pole of a magnet near the *N* pole of a second magnet which is suspended. It is found that there is a repulsion between these ends of the two magnets. If on the other hand, the *N* pole is brought near the *S* pole of the suspended magnet, there is found to be an attraction between the two magnets.

Hence we have the law: *Like-named magnetic poles repel each other, and unlike-named magnetic poles attract each other.*

The above law gives us a means of testing whether a bar of iron or steel is a magnet or simply *a magnetic substance*, that is, a substance attracted by a magnet. If a steel bar shows at any point a repulsion for the *N* pole of a suspended magnet, the bar is a magnet and the repulsion indicates the location of the *N* pole of the bar. If the steel or iron bar is not magnetized, every point of the bar shows attraction for either pole of the magnet.

366. Experiment of Breaking a Magnet.—When a strongly magnetized needle is dipped into iron filings it is found that the

filings cling in tufts at the poles near the ends, but that there are no filings near the middle, or the neutral region. If now we break the needle at the middle, we get two complete magnets. Upon testing each half, we find that a S pole appears on the side of the break towards the original N pole, and a N pole on the side towards the original S pole (Fig. 230). Each half can be in turn broken, and four magnets obtained and so on indefinitely. We are thus lead to consider that a piece of iron or steel is made up of elementary magnets with equal and opposite poles and that the magnetization of a bar of iron or steel consists in arranging these elementary magnets of the bar. In an unmagnetized steel or iron bar, there is no general trend of these elementary magnets in any one direction, and so they neutralize each other's external action (Fig. 231); if, however, we can by any means turn the majority of the elementary magnets of a bar in the same general direction, then the bar becomes a magnet (Fig. 232). The above explanation of the action of a magnet has been termed the *molecular theory of magnetism*, but it is not a necessary part of the theory that the elementary magnets are molecules in the chemical sense (see §496 on Electron Theory of Magnetism).

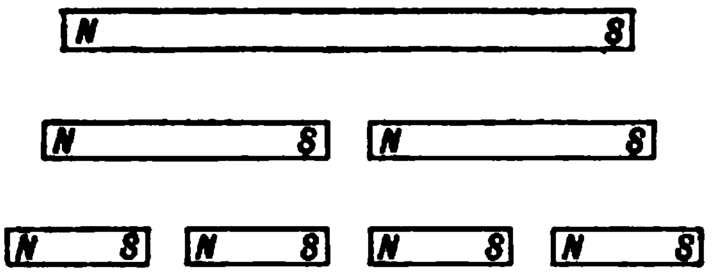


FIG. 230.

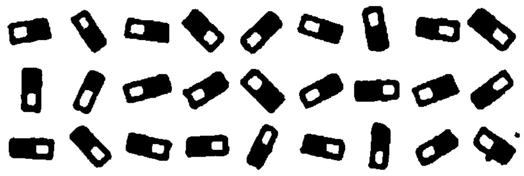


FIG. 231.

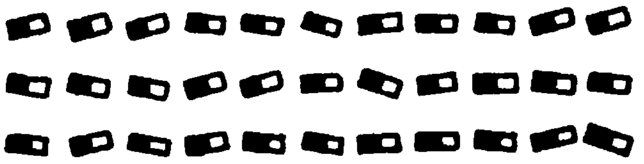


FIG. 232.

An experiment to illustrate the above theory of magnetism is the magnetization of steel filings contained in a glass tube. By stroking the tube on a strong magnet the filings, which in general are magnets, may be lined up in the direction of the tube, so that there results a N pole at one end and a S pole at the other end. This is shown by bringing the tube up to a delicately suspended magnetic needle. If now the filings are shaken up, so that the small magnets are no longer lined up in any particular direction, it is found that either end of the tube of filings attracts either pole of the suspended needle, that is, the tube of filings has lost its magnetization.

Ewing's model of a magnet consists of a number of small magnetic needles, mounted on pivots upon a glass plate, (Fig. 233). The magnetic needles take no special direction unless acted on by one or more large steel magnets or by the directive magnetic action of a coil carrying an electric current (see §427). Ewing's model can be made of a size to be put in the vertical beam of a projection lantern, and thus the motions and directions of the small magnets can be clearly shown to a large class. This model is also used to explain properties of a magnet which depend upon the mutual action of the elementary magnets (see Hysteresis, §497).

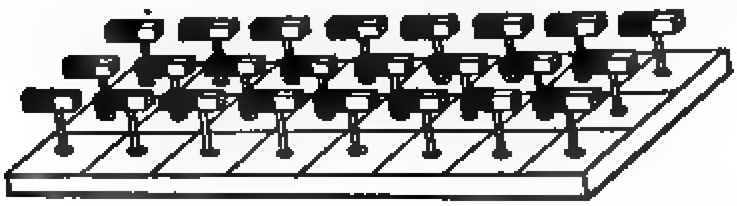


Fig. 233.

367. Magnetic Lines of Force, Magnetic Field.—We have seen that a magnet acts upon a neighboring magnet, the unlike poles attracting and the like poles repelling each other. This action was assumed by earlier writers as "direct action at a distance," that is, as taking place across space and not by means of any intermediate actions. Faraday and Maxwell showed that the action of one magnet upon another is to be explained

Fig. 234a.

as due to lines of stress which exist in the space between the magnetic poles. These lines of stress can be traced by methods described below, and are called "lines of magnetic force."

Fig. 234b.

Fig. 234c.

Magnetic force as transmitted along these lines may be thought of as similar to a "pull" along a cord. We shall see later that

a magnetic line of force is probably the line of the centers of whirls in the intervening space.[1]

The lines of force between magnetic poles can be traced by means of iron filings. Thus, Fig. 234*a* shows the tracings made by filings of the magnetic lines between a *N* and a *S* pole. It is formed by sprinkling iron filings on a glass plate, with the two poles beneath the glass, and at the same time gently tapping

FIG. 234*A*.

the glass so that the filings are free to move. The filings arrange themselves in lines which show the magnetic lines of stress. Each particle in the filings becomes, for the time being, a magnet, the *N* pole of which tends to move in one direction along a line of force, while the *S* pole tends to move in the opposite direction. Fig. 234*b* shows the lines of force as traced by lines about a bar magnet. In Fig. 234*c* we have the lines between two *N* poles; in Fig. 234*d* the lines around a soft iron bar in the field between a *N* and a *S* pole; in Fig. 234*e* the field around a soft iron ring placed between a *N* and a *S* pole. In the last figure, it is seen that the filings form no lines inside of the ring, that is, that region is shielded from the magnetic force (see §492).

[1] Faraday and Maxwell, and most students of physics, have explained the transmission of magnetic, electrical and luminous effects, by assuming the existence of a medium called the "ether." The conception of the "ether," has been one of the most helpful and convenient theories in science, but it has never been without difficulties. To explain all the observed facts of magnetism, electricity and light requires us to assume a medium which has "properties" which are difficult to reconcile with each other. The "ether" is, however, the best working hypothesis which we have for these phenomena and as such we use it in the discussion of electricity and magnetism.

368. Magnetic Field.—A region in which lines of magnetic force exist is called a *magnetic field*. All the region about a magnet is thus a magnetic field. It will be shown later that the region about an electric current is also a magnetic field. The earth is surrounded by a magnetic field, known as the earth's magnetic field. A sensitive test of a magnetic field is the exertion of force on a delicately suspended magnetic needle. Such a magnetic needle is acted on by fields which are too weak to turn iron filings. Thus the earth's field does not rotate iron filings, but it acts on a suspended magnetic needle.

When the magnetic lines in a field are parallel to each other, the field is a *uniform field*. The earth's magnetic field, in places

Fig. 234a.

free from masses of magnetic substances and distant from electric currents, is practically uniform over considerable areas. A suspended magnetic needle points in practically the same direction throughout such a field.

In mapping a field by a magnetic needle, we note that the suspended needle places itself tangentially to the magnetic line through its center. The positive direction of the magnetic line is that in which the N pole of the needle tends to move. It is thus seen that a magnetic line in air starts from the N pole of a magnet and ends in a S pole. But we have seen that a N and a

S pole attract each other; that is, the magnetic lines tend to shorten or contract. Indeed, in Faraday's thought, the attraction between the two unlike poles is due to the tension of the lines which join the poles, these magnetic lines acting like stretched rubber cords. It is noted from the tracings of the lines that magnetic lines are in general curves. Faraday, in fact, often referred to them as "magnetic curves." If, however, the only property of a magnetic line were that of contraction, the line would be straight. But it is to be noted that lines diverge from each other; that is, the general form of the lines seems to be due to two forces, (*a*) a tension along the line, and (*b*) a repulsion between the lines, the last acting like a pressure at right angles to the lines. James Clerk Maxwell has shown mathematically that the properties of a magnetic field and the resulting forces acting on magnets can be accounted for completely by the longitudinal stress and the lateral or perpendicular pressure in the medium.

369. Methods of Magnetization.—We can magnetize a rod of soft Norway iron by simply holding it in a vertical plane through the north-and-south line and inclined downward about 70° from the horizontal; after a very little tapping or perhaps none, it is found that the end pointing northward has become a N pole. That is, the elementary magnets of the rod have been lined up under the action of the earth's magnetic field. Upon placing the rod at right angles to the earth's magnetic field and again tapping it lightly, it is found to have lost its magnetization as readily as it gained it. The rod is now easily magnetized in the opposite direction by reversing it from its first position, and gently tapping it. If we try the same experiment with a piece of hard iron or of tool steel, it is found that the hard iron or steel can be magnetized in the earth's field only by sharp and prolonged tapping. It is also found that when the hard iron and steel rods are once magnetized, they retain their magnetization, even when their position in the field is changed. This property of retaining magnetization is called *magnetic retentivity*. (The term *coercive force*, once used for *magnetic retentivity*, is now used in a different sense. See §497 on Hysteresis, etc.). The elementary magnets of soft iron are thus easily lined up, but are as easily thrown out of line again. The elementary magnets of steel

resist a change of direction, and hence the steel is less easily magnetized, but when once it is magnetized, it retains its magnetization. Tool steel is accordingly adapted for *permanent* magnets; soft iron, only for *temporary* magnets. It is found that the retentivity of steel is greatly increased by tempering it, so that strong permanent magnets are always made of steel tempered hard.

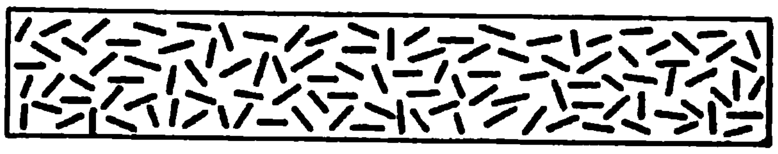


FIG. 235.

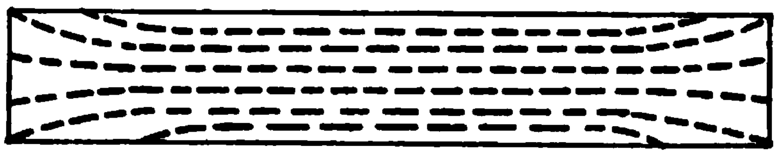


FIG. 236.

From the above we see that the process of magnetization consists in bringing the iron or steel into a magnetic field. Figs. 235 and 236 illustrate what takes place according to the molecular theory. In the first figure, the small steel magnets point in all directions, and the lines of force are practically all inside the group of magnets. In the second figure, the small magnets are pointed in one general direction, and the external field is approximately that of a bar magnet.

To magnetize an iron or steel bar so that it is a strong magnet, it is necessary to line up a large part of the elementary magnets of the bar, and this calls for a strong magnetic field. Hence, such a weak magnetic field as that of the earth gives only comparatively small magnetizing effects. Strong magnetic fields are obtained by using strong steel magnets, or strong electromagnets (§483) or solenoids with a large number of ampere-turns (§430). We shall study later (§484) the quantitative relations between the strength of the magnetic field and resultant intensity of magnetization for various kinds of iron and steel (§486).

370. Magnetic Substances and Induced Magnetism.—If a piece of soft iron or steel such as a nail is brought near the *N* pole of a strong magnet, not only is it attracted, but it acquires the property of attracting other nails; thus a whole series or chain of nails may be held up by the poles "induced" from nail to nail, each nail becoming for the time a magnet (Fig. 237). That is, the attraction is really an attraction between the *N* pole of the magnet and the *S* pole that is induced in the nail. Magnetic

substances thus are substances which become magnetic by induction and hence are attracted by a magnet. Iron, and to a less degree, nickel and cobalt and an alloy of copper, manganese and aluminum, called the "Heusler alloy," are the substances showing magnetic properties most strongly, and are called *ferromagnetic*, but many other substances show a slight magnetic attraction in very strong fields. Such substances are called *paramagnetic*. Still other substances as bismuth, are repelled by a strong magnet. Such substances are called *diamagnetic*. The quantitative relations of the magnetic properties of substances will be discussed under Magnetic Induction (§485).

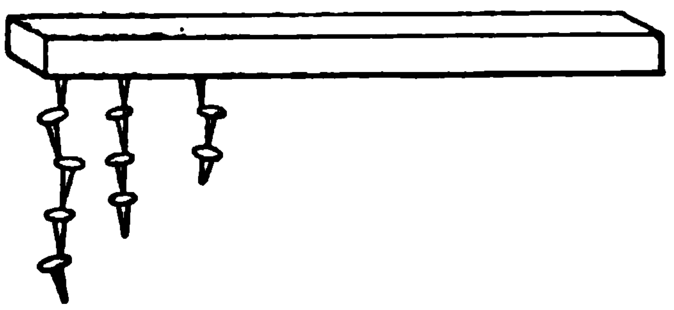


FIG. 237.

371. Intermediate Poles.—As "normally" magnetized a needle has only two magnetic poles, but it is possible to magnetize a steel needle so as to have more than two points of maximum magnetic attraction (Fig. 238*b*). In Fig. 238*a* is shown one method of securing such an irregular magnetization. The bar is placed so that each end rests on the *N* pole of a bar magne and it is stroked at the middle with the *S* pole of a third magnet. The bar will then be found to be magnetized with a *S* pole at each end and a *N* pole in the middle It is evident that we have in this case the equivalent of two magnets with the two *N* poles at the middle. The intermediate pole is sometimes called a "*consequent*" *pole*.

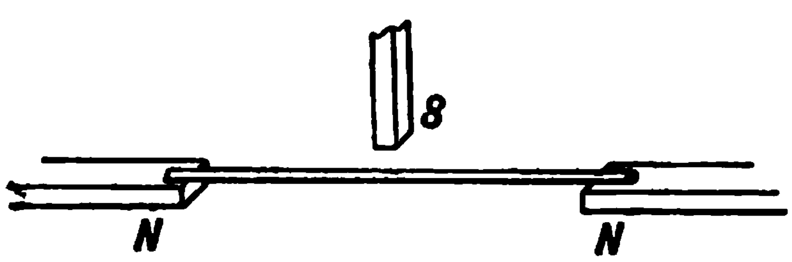


FIG. 238*a*.

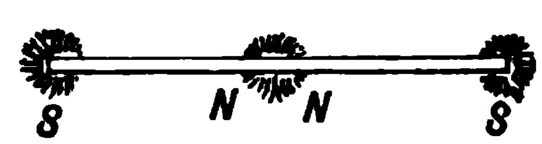


FIG. 238*b*.

372. Coulomb's Law of Magnetic Force.—In the case of a slender knitting needle, which has been magnetized in a strong uniform magnetic field, the elementary magnets are generally so perfectly in line that a magnetic pole of the magnetized needle can be assumed for distances greater than a few centimeters to be a point very near the end of the needle. (See Fig. 229.) Coulomb, a French physicist and mathematician, in 1789 used slender magnetic needles to study the force beween two magnetic poles when placed at different distances. He found that the

force between the poles varied inversely as the square of the distance between the poles. Coulomb's method of experimenting with "the torsion balance" can be represented diagrammatically as follows: A long and thin needle NS (Fig. 239) is suspended horizontally by a thin silver wire. This suspension wire is free from torsion when the needle is in the magnetic meridian. A second slender needle $N'S'$ held vertically is brought so that the horizontal distance between the two north poles is d (as measured before any deflection of NS is allowed). If NS is free to move, it is deflected by the repulsion between the two N poles. To bring NS to its original position a twist must be given to the suspension wire, by turning the torsion head until the force of torsion is equal to the force of repulsion between the two magnetic poles. The force between the two poles is measured by the number of degrees of torsion in the wire (§119). By thus measuring the forces F', F'', F''', etc., for the distances d_1, d_2, d_3, etc., between the two poles, Coulomb was able to show that $F' : F'' : F''' :: 1/d^2_1 : 1/d^2_2 : 1/d^2_3$, that is, that F is proportional to $1/d^2$.

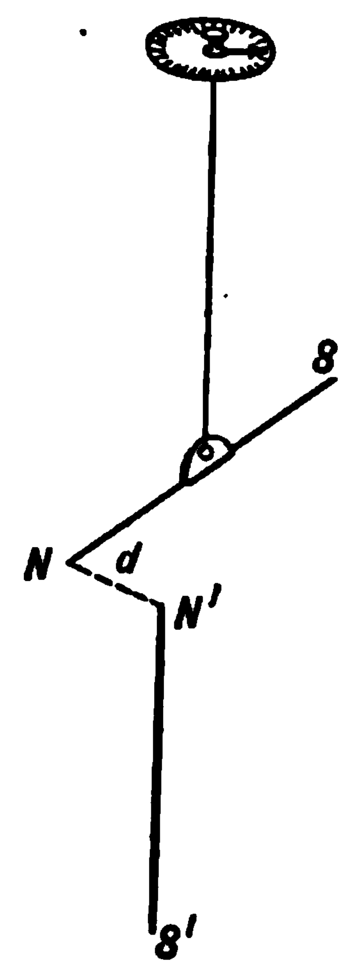


FIG. 239

We can now define the c. g. s. *unit magnetic pole*, or pole of unit strength; *a unit magnetic pole is one which when placed at one centimeter distance in a vacuum from an equal and like pole repels it with a force of one dyne.* Hence, if the centimeter be used as the unit of distance, the dyne as the unit of force, and we measure m and m' in terms of the above c.g.s. unit of pole strength, Coulomb's law for a vacuum becomes

$$F = \frac{mm'}{d^2}.$$

If instead of a vacuum there is a material intervening medium, a factor for that medium $1/\mu$ (§490) must be used, and we have

$$F = \frac{1}{\mu}\frac{mm'}{d^2}$$

The factor $1/\mu$ is for all practical purposes equal to unity for air. The proof of Coulomb's law does not rest upon Coulomb's

experiment, which is necessarily approximate, but upon the fact that the action of magnets on each other in various positions can be predicted by the use of Coulomb's law (see §377).

373. Intensity or Strength of a Magnetic Field.—The force F which acts on a magnetic pole placed in a magnetic field depends upon (*a*) the strength m of the pole and (*b*) on what may be called the strength H of the field. This suggests the following definition of the strength or intensity of a magnetic field: *The strength or intensity of a magnetic field at a point is equal to the number of dynes of force which act on a unit magnetic pole at the point.* Hence $F = m \times H$. From this formula we can calculate the force acting on a magnetic pole if we know the pole strength and the field intensity, it being assumed that the strength of the magnetic field is not appreciably changed by the presence of the testing pole. Thus in the earth's field of intensity 0.6 a pole of strength 4, is acted on with a force of $0.6 \times 4 = 2.4$ dynes.

The unit field intensity is by some writers called the *gauss*. Thus the earth's magnetic field at Washington would be described as a field of 0.6 gausses.

374. Quantitative Use of Lines of Force.—Magnetic lines of force, as they have been defined above (§367), fix only the direction of the field. The fact that in the figures made by iron filings, the lines appear most numerous where the field is strongest, suggests that the intensity of the field may be represented by the number of the lines of force. To this end, we agree to restrict the number of lines drawn to represent a magnetic field so that in a field of unit intensity there is one line of force per square centimeter of normal section, and in a field of intensity H there are H lines per square centimeter; or *the intensity of the field*, as defined above, *is numerically equal to the number of lines per square centimeter* cutting a plane at right angles to the field. That a line is continuous in the field will appear from a consideration of the lines entering and leaving poles, and also from the very nature of a line of force.

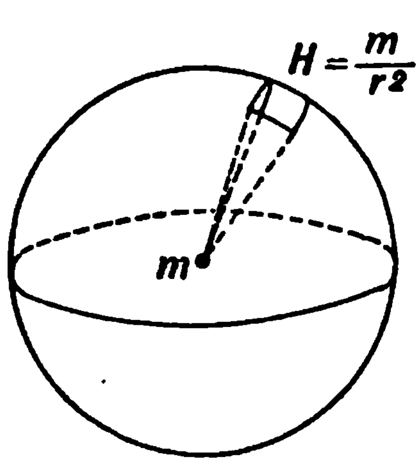


FIG. 240.

If the field is uniform, the total number of lines across the

field of section S would thus be $N=SH$. If the field is not of uniform intensity,

$$N=(S_1H_1+S_2H_2+, \text{etc.})=\Sigma SH$$

By the following consideration we see that $4\pi m$ *lines of force emerge from a pole* $+m$ *in a vacuum.* Describe about the pole m as center a spherical surface with radius r (Fig. 240). The intensity of the field on the sphere is by Coulomb's law

$$H=\frac{m\times 1}{r^2}.$$

This field H is evidently at right angles to the surface of the sphere. Hence there are m/r^2 lines across each square centimeter of area of the sphere. The total number of lines coming from the pole is therefore

$$N=SH=4\pi r^2\times m/r^2=4\pi m$$

Since this is true for any and every value of r, these lines are continuous. In a similar way we find that $4\pi m$ lines *enter* the pole $-m$.

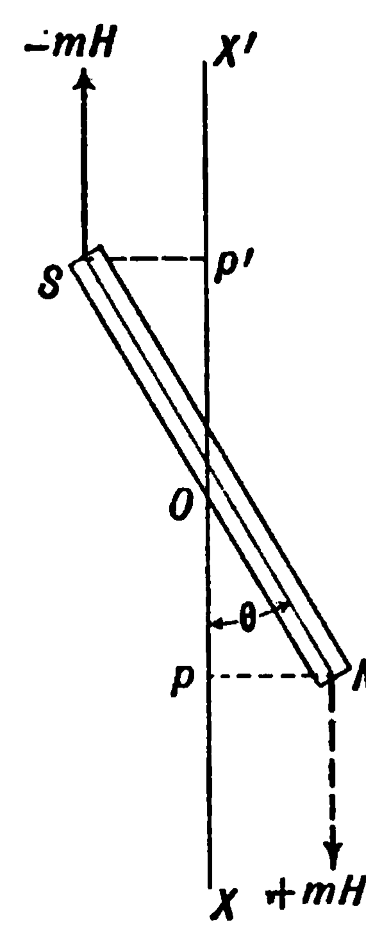


Fig. 241.

375. Forces on a Magnet in a Magnetic Field.—A magnetic needle in a uniform magnetic field, such as the earth's field, is acted on by two equal and opposite forces, the force $+mH$ on the N pole, and the force $-mH$ on the S pole, Fig. 241. We thus have two equal and opposite forces acting at opposite ends of the magnet, that is, we have a couple (§98). The action on the magnetic needle is simply to rotate it into the line of the field, without translation. This can be easily verified by floating a magnetic needle on a cork in a large basin of water. The needle is not drawn to the north nor to the south but simply rotates and finally comes to rest in the magnetic meridian. We can also regard the experimental fact that there is no tractive force on a magnet in the uniform field of the earth, as a proof of the assumption made above, that the two poles $+m$ and $-m$, are equal in strength.

If the field is not uniform, but is stronger at one pole than at the other, there will be a tractive force on the magnet. It is

then not only rotated, but also acted on by a resultant force in one direction or the other.

376. Torque in a Uniform Field.—The torque or moment of the couple acting on a magnetic needle in a uniform field is easily expressed. Consider the magnet NS of length l, making the angle θ with the direction of the field XX' (Fig. 241). If $+m$ and $-m$ are the strengths of the poles, and H the intensity of the field, the field exerts two parallel forces $+mH$ and $-mH$, and the moment of the couple is $L=mH\times$ (arm of the couple). The arm of the couple $=2Np=2ON \sin\theta=l\sin\theta$.

Hence $\qquad L=Hml\sin\theta.$

Put $ml=M$, then $\qquad L=HM\sin\theta.$

The term ml is called the *magnetic moment of the magnet. The magnetic moment of a magnet is equal to the product of the pole strength of the magnet by the distance between the poles.* When the magnet is held at right angles to the field, that is, when $\theta=90°$, the torque is $L=HM$.

If the field strength H is unity, then the torque L is equal to the magnetic moment M, and it follows that: *The magnetic moment of a magnet is numerically equal to the torque acting on it when it is held at right angles to unit field.*

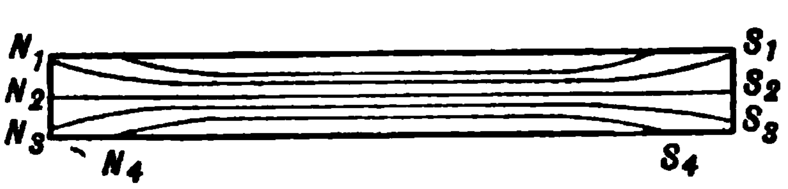


FIG. 242.

It is to be noted that the magnetic moment of a magnet is a quantity that admits of exact determination, while the strength of the pole m, and the distance l between the two poles, cannot be exactly determined in a physical magnet.

It is easy to see that the magnetic moment of a bundle of magnets is equal to the algebraic sum of the magnetic moments of the individual magnets. But a physical magnet is to be looked upon as a bundle of magnetic filaments. By a magnetic filament we mean a single line of elementary magnets, arranged as in Fig. 242, with only two poles "free," that is, poles not neutralized by the presence of equal opposite poles. These two "free" poles N and S, form the poles of the filament. Hence the magnetic moment of the whole magnet is the algebraic sum of the magnetic moments of these filaments of the elementary magnets.

377. Calculation of the Intensity of the Magnetic Field in Special Cases.—It is possible by the use of Coulomb's law to calculate the strength of the magnetic field at certain points about a magnet of known magnetic moment. The cases of most importance are for the two positions known as "position A," and "position B of Gauss."

For Position "A."—Consider the strength of the magnetic field, due to a bar magnet at a "distant" point on the line of its axis. The strength of the pole is m, the distance between the poles or length of the magnet is $2L$; the problem is to find the strength of the field at a point P in the line of the axis, and distant r from the mid-point of the axis of the magnet (Fig. 243). By Coulomb's law, the force on a unit positive pole at P is

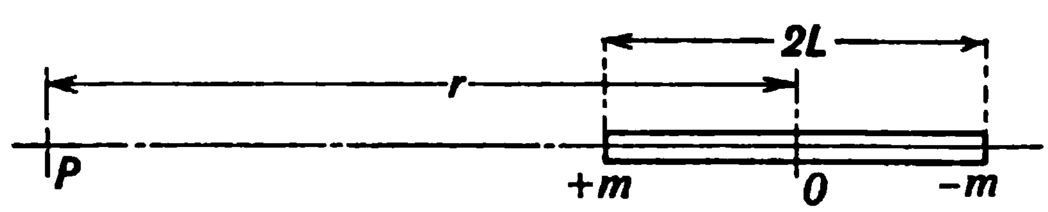


FIG. 243.

$$F=\frac{m}{(r-L)^2}-\frac{m}{(r+L)^2}$$

$$=\frac{4rLm}{(r^2-L^2)^2}$$

If P is "distant," so that L^2 can be neglected as compared with r^2, we get

$$F=\frac{4Lm}{r^3}=\frac{2M}{r^3};$$

or the strength of field at P is

$$H_P=\frac{2M}{r^3}. \qquad \text{(A)}$$

The direction of this field is evidently that of the axial line OP.

For Position "B."—Consider the strength of field due to a magnet at a "distant" point on the line bisecting the axis at right angles (Fig. 244).

In this case the forces acting on unit pole at P, are a repulsion due to $+m$, represented by PA, and an attraction due to $-m$, represented by PB. The resultant is represented by the diag-

onal PR. Since the triangles PAR and NPS are similar, we have,

$$\frac{PR}{PA}=\frac{NS}{NP}=\frac{2L}{(r^2+L^2)^{1/2}}$$

But PA represents the force exerted by $+m$ on the unit positive pole at P, or from the law of Coulomb

$$PA=\frac{m}{(r^2+L^2)};$$

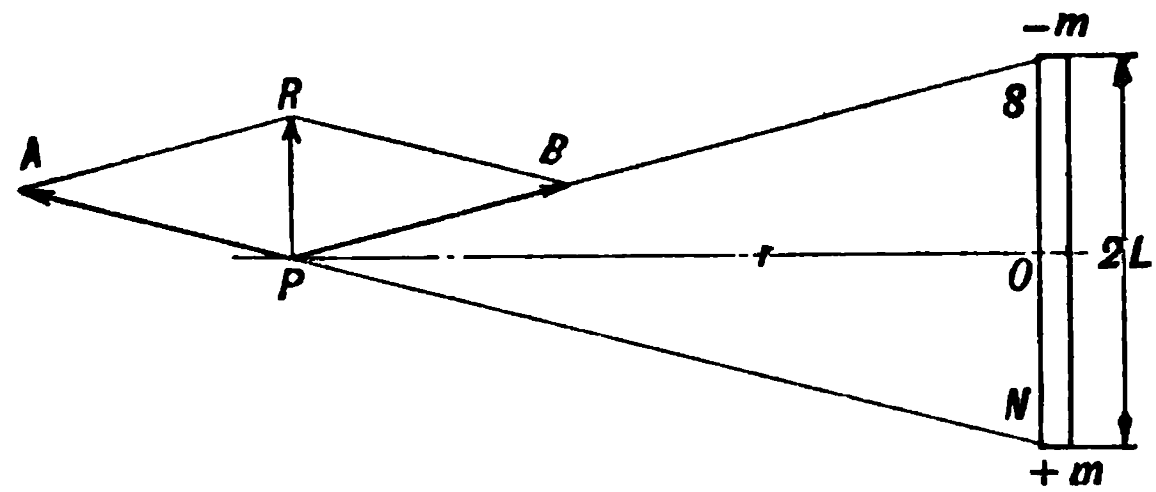


FIG. 244.

substituting this value for PA, and transposing we get the resultant force

$$H_P=PR=\frac{2Lm}{(r^2+L^2)^{3/2}}$$

If P is "distant," so that L^2 can be neglected as compared with r^2, we get

$$H_P=\frac{2Lm}{r^3}=\frac{M}{r^3} \qquad (B)$$

The direction of this field is evidently perpendicular to the bisecting line OP, or parallel to the magnet.

From the above it is seen that the intensity of the field for position "A" is twice that for position "B" for the same magnet and the same distance. As the calculations have been made on the assumption of Coulomb's law, we have here a means of testing this law by comparing experimentally the intensities of the fields in the two cases. This has been done by Gauss and the results of experiments agree with the law.

378. Methods of Comparing the Intensities of Two Fields.—Since the force acting on a magnetic pole is proportional to the intensity of the field, (that is, $F=mH$) the ratio of the intensities

of two fields is equal to the ratios of the forces which act on the same magnetic pole in the two fields; thus $H_1 : H_2 :: F_1 : F_2$, where H_1 and H_2 are the intensities of the two fields, and F_1 and F_2 are the two forces on the pole m in these fields. The forces F_1 and F_2 can be measured by the following methods:

(a) By balancing the torque on a suspended magnet by the torsion of a suspension wire.

(b) By the vibrations of an oscillating magnet.

(c) By the deflections produced by a second magnet.

379. Comparison of Magnetic Fields by the Torsion Balance.—First suspend the magnet by a wire (or quartz fiber) suspension, and arrange so that there is no torsion in the suspension wire when the needle is in the direction of the field. Next twist the wire by means of the "torsion head" until the needle is deflected through a given angle ϕ The number of degrees of torsion, x, in the wire, is calculated from the reading of the torsion head and the deflection of the magnet. Then $x_1 = kMH_1 \sin \phi_1$ (§376) where k is a constant for a given suspending wire (§168). If we repeat this experiment with the same magnet and the same suspension arrangements in a second magnetic field we find a torsion x_2 for deflected ϕ_2. Thus $x_2 = kMH_2 \sin \phi_2$. From this it follows directly that

$$H_1 : H_2 = x_1/\sin \phi_1 : x_2/\sin \phi_2$$

If the deflection ϕ_1 be made equal to ϕ_2, the proportion becomes

$$H_1/H_2 = x_1/x_2$$

380. Comparison of Fields by the Oscillations of a Magnet.—When a suspended magnet is deflected through an angle θ from the direction of the field, it is acted on by a restoring couple $MH \sin \theta$ (§376). For small angles, the sine and the angle are assumed equal, and hence the restoring couple is proportional to θ, and $MH\,\theta = -I\alpha$, where I is the moment of inertia, and α the angular acceleration (see §89). Hence the motion agrees with the definition of angular harmonic motion (§118), and the period T is given by the formula

$$T = 2\pi\sqrt{I/HM}$$

Transposing we get,

$$H = \frac{4\pi^2 I}{M}\,\frac{1}{T^2} = \frac{4\pi^2 I}{M} n^2$$

where n is the frequency of the vibration. By allowing the same needle to vibrate in two fields of strength H_1 and H_2, and noting

the periods T_1 and T_2 or frequencies n_1 and n_2, we get the proportion,

$$H_1 : H_2 = 1/T_1^2 : 1/T_2^2 = n_1^2 : n_2^2$$

In the above, it is assumed that the moment of the magnet is constant, and, therefore, the method cannot be used in strong fields, that is, in fields which change the moment of the magnet by induction. Fig. 245 shows simple apparatus for making observations of the oscillations of a magnet. The magnet is a cylinder, the moment of inertia of which can be calculated by formula (§93).

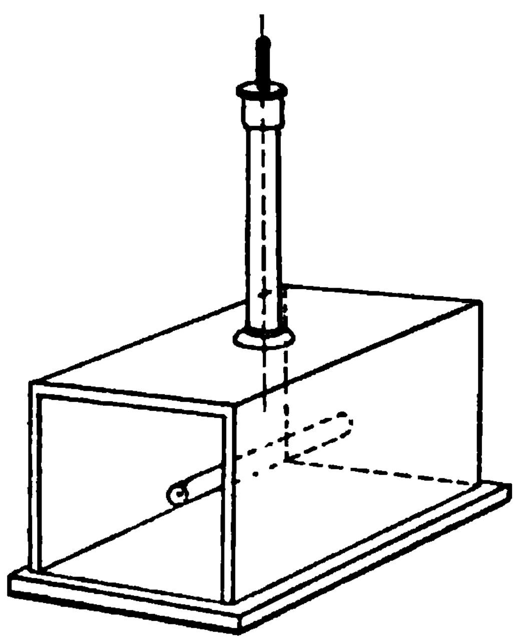


FIG. 245.

381. The Tangent Law.—When a magnet is under the action of two fields H and R, which are at right angles to each other, it takes a resultant position making an angle θ with H, and an angle $(90°-\theta)$ with R, (Fig. 246). The moment of the couple tending to rotate it into the direction of H is $L_1 = MH \sin\theta$, and that into the direction of R is $L_2 = MR \sin(90°-\theta) = MR \cos\theta$. Since the magnet comes to rest at the deflection θ, the two opposite torques L_1 and L_2 must be numerically equal, that is, $L_1 = L_2$ or

$$MH \sin\theta = MR \cos\theta$$

From this we get

$$\frac{R}{H} = \frac{\sin\theta}{\cos\theta} = \tan\theta$$

Hence: *If a magnetic needle in a field of intensity H is deflected through an angle θ by a field R at right angles to H, then the tangent of the angle of deflection θ is equal to the ratio of the strengths of the two fields R and H.*

The tangent law is used in the tangent galvanometer (§436), and most other magnetic deflection instruments. It is an application of the general law that the ratio of two rectangular forces is equal to the tangent of the angle which the resultant makes with the first component.

382. Comparison of Magnetic Fields by the Deflection Experiment.—In this method a small magnetic needle is deflected from

the direction of the field by a second magnet, which is placed so as to produce a field at right angles to the field to be measured. A simple form of apparatus for this experiment is shown in Fig. 247 *a* and *b*. It consists of a magnetic compass O

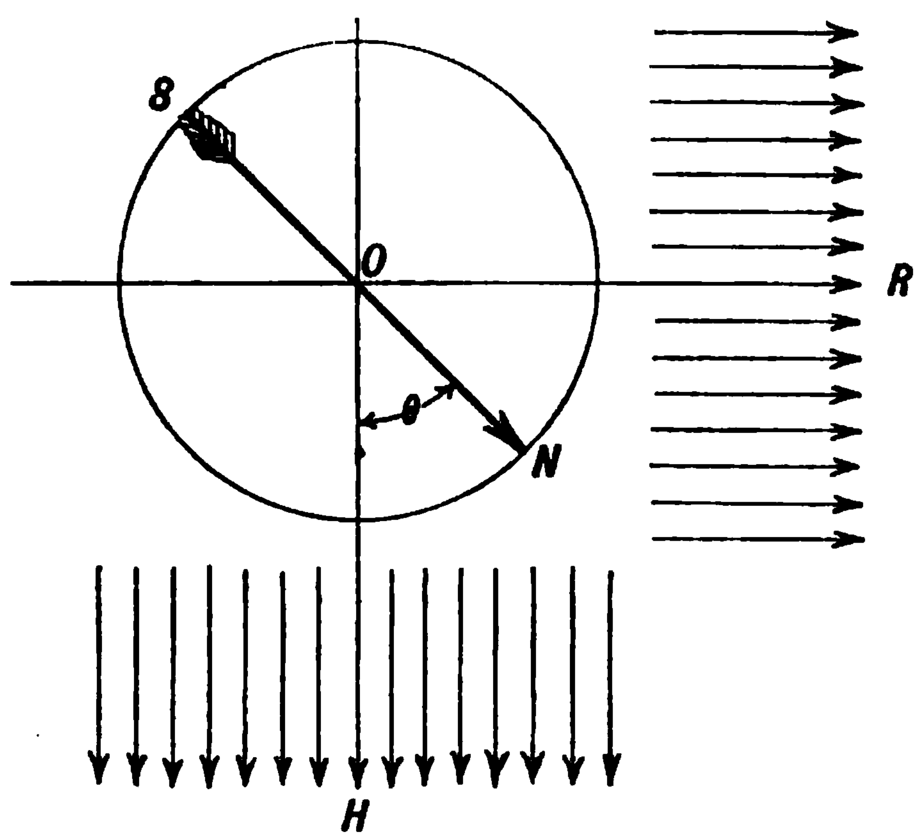


FIG. 246.

mounted in the middle of a graduated bar AB. The compass box is arranged with graduated circle so that the deflection of the needle can be read. The bar AB is set at right angles to the magnetic field H_1, and the zero position of the needle is read on the graduated circle. A magnet NS, is now placed at a point

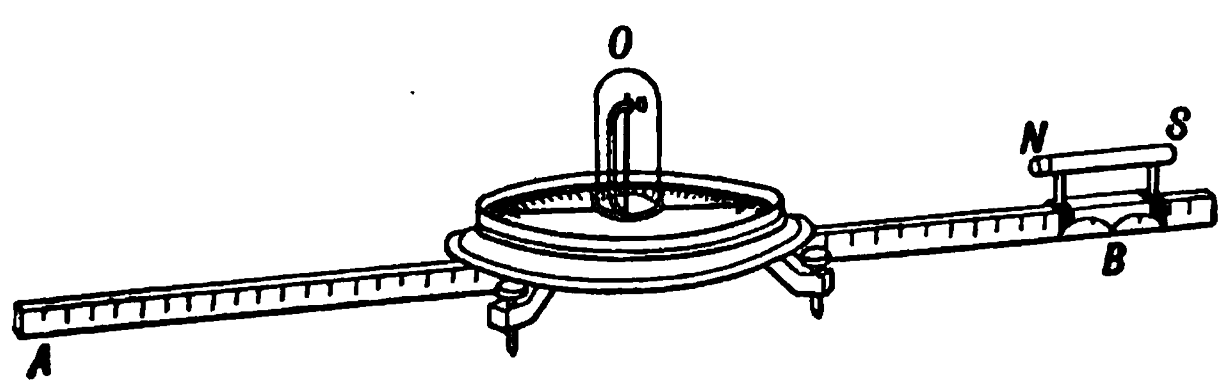


FIG. 247*a*.

B on the bar, and it then produces a magnetic field of strength R at right angles to the field H_1. The needle takes a resultant position, making an angle θ_1 with the field H_1, such that $R/H_1 = \tan\theta_1$ (see §381). We can now transfer the apparatus into a second magnetic field H_2, and get a second angle of de-

flection θ_2. Then $R/H_2 = \tan\theta_2$. Dividing the second equation by the first, we get,

$$\frac{H_1}{H_2} = \frac{\tan\theta_2}{\tan\theta_1}$$

The ratio of the tangents of the two angles of deflection thus gives the ratio of the intensities of the two magnetic fields.

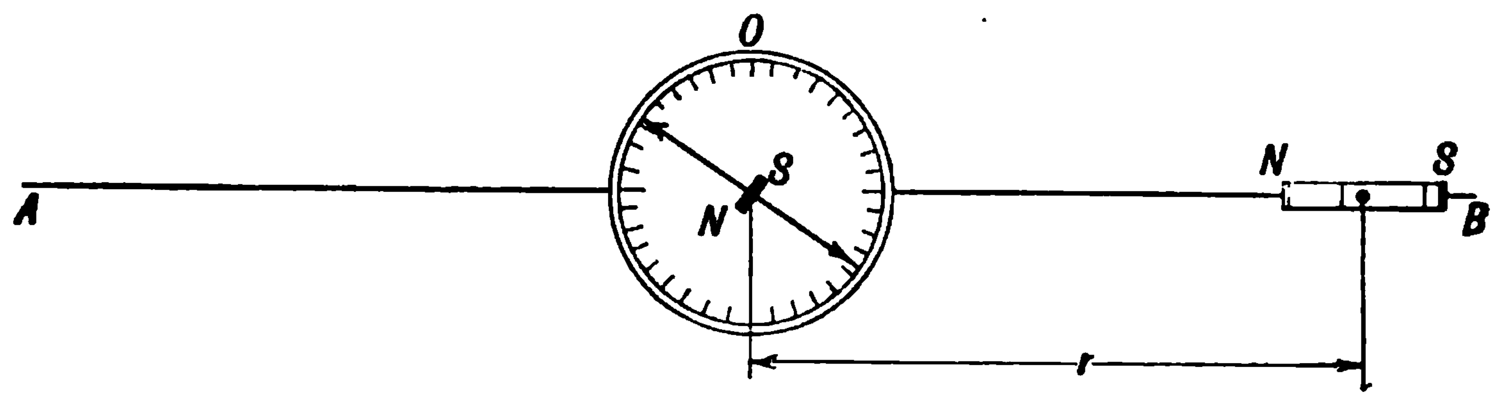


FIG. 247b.

383. Determination of H and M in Absolute Measurements.—The methods described above (§§380, 382) are comparison methods, that is, the relative strengths of magnetic fields are determined, but not their absolute values. The absolute measurements of a magnetic field such as that of the earth can be made by a combination of the oscillation and the deflection experiments as follows:

First, a magnet of moment M is allowed to vibrate freely in a field of strength H, and we thus get the relation (§380),

$$HM = \frac{4\pi^2 I}{T^2} \tag{I}$$

By this, the value of the product HM is obtained.

Second, using the same magnet of moment M as the deflecting magnet in the deflection experiment, (§381), we get

$$R/H = \tan\theta$$

But in §377, we have seen that the field due to a magnet in the "A" position is $R = 2M/r^3$. Substituting this value, we get

$$\frac{H}{M} = \frac{2}{r^3 \tan\theta} \tag{II}$$

where r is the distance in centimeters from the center of the magnet M to the center of the deflected magnet (Fig. 247), θ is the angle that the deflected magnet makes with the field H, and M is the magnetic moment of the deflecting magnet. Combining equation (I) and (II), we can eliminate M, and get H in terms of observed quantities and numbers, that is, in absolute measure.

(This "strength of field" is the horizontal component of the earth's field, see §387.)

In a similar way we can get M, the magnetic moment of the magnet, in absolute measure by eliminating H between the equations (I) and (II).

384. Magnetometers to Determine the Horizontal Component of the Earth's Magnetic Field.—By the use of simple apparatus such as shown in Figs. 245 and 247, the value of H, the horizontal component of the earth's magnetic field, can be determined to an accuracy of a few per cent. For the most accurate work, such as is required in the magnetic surveys of the governments of the United States and Great Britain, the "Kew" unifilar magnetometer is used. This is shown in Fig. 248 arranged for deflection experiments. The general method is that of the simpler apparatus, but special details and corrections are involved for which the larger laboratory manuals must be consulted.

385. The Earth a Magnet.—The fact that a suspended magnetic needle tends to place itself in a north-and-south line, led to the theory that "the globe of the earth is a great lodestone," and that the positive magnetic pole of the earth is near its south geographical pole, and its negative magnetic pole is near its geographical north pole. Sir William Gilbert, rightly called "the father" of magnetism as a science, first published this theory in 1600 in his famous book the "De Magnete." But later study has shown that the magnetization of the earth is very complex and the two so-called "magnetic poles" of the earth, in the northern and southern hemispheres respectively, must not be regarded as closely analogous to the pole of an ordinary magnet, but are merely places where the magnetic force is perpendicular to the earth's surface.

The study of the earth's magnetic field is one of the most important and interesting fields of science, because it involves the problem of how and why the earth is a magnet, and also because of the use of the magnetic compass in navigation and surveying and in the absolute electrical measurements. A complete description of the earth's magnetic field calls for determinations of (*a*) the direction, and (*b*) the intensity of the field for every part of the earth.

386. Direction of the Earth's Magnetic Field. Declination, Dip.—It is found that a magnetic needle which is suspended so as to rotate in a horizontal plane, does not in general point exactly to the geographical north. The angle which such a needle makes with the geographical meridian is called *the declination*. This angle varies with both place and time. Thus, in 1905, the declination of the magnetic needle at London was 16° 32.9′ W of

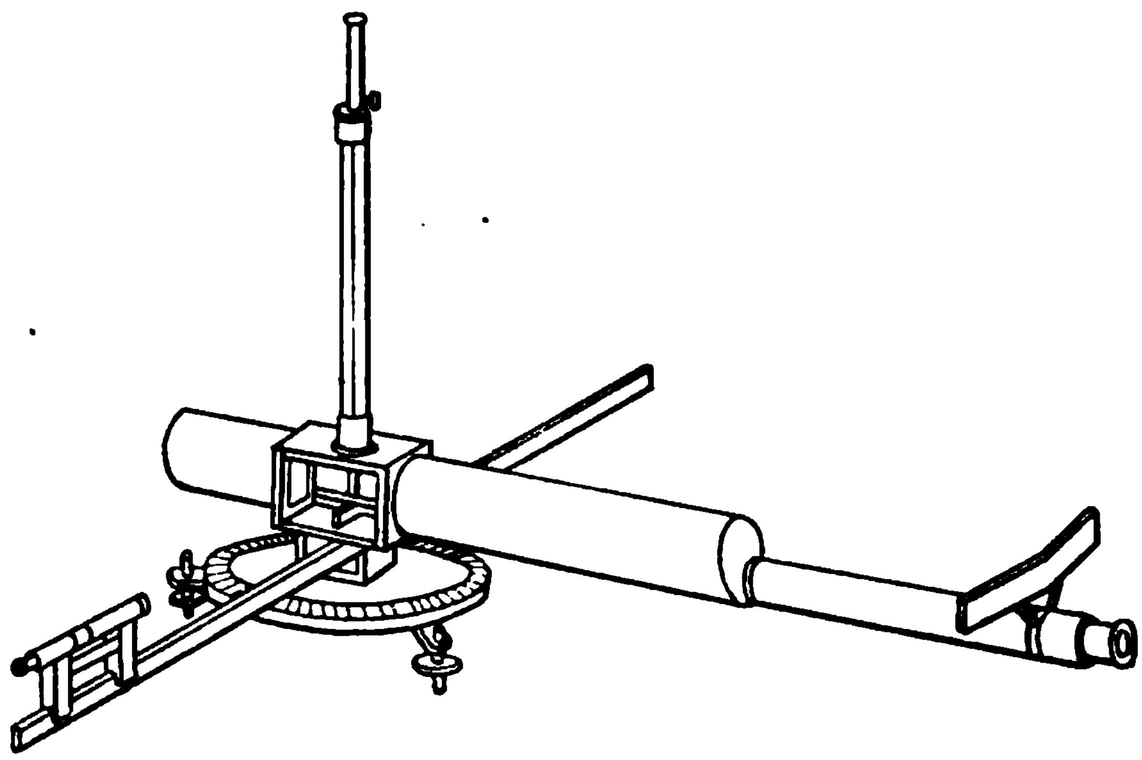


FIG. 248.

north; at New York it was 9° 0.8′ W, and at San Francisco 16° 55′ E. There are also variations with time, but these are generally slow or transient and will be considered in §389. A vertical plane through the axis of a compass needle intersects the earth in a line called the *magnetic meridian*. Evidently the declination at any place can be defined as the angle between the geographical and magnetic meridians.

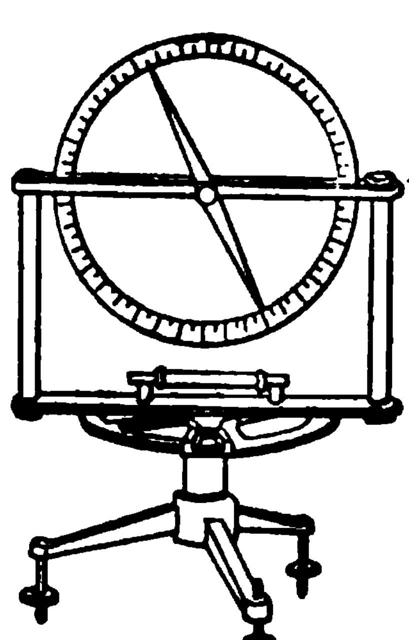


FIG. 249.

About 1544, Hartmann observed that a needle which was balanced horizontally when non-magnetized, was no longer balanced when magnetized, but dipped with its north pole downward. The *magnetic dip or inclination* thus observed can be measured by an instrument called a *dip circle*. This consists (Fig. 249) of a vertical graduated circle which can be set in the magnetic meridian. At the center of the circle is a magnetic needle which is balanced on a horizontal axis through its center of gravity, so as to rotate freely

in the plane of the magnetic meridian. The angle which a magnetic needle balanced at its center of gravity makes with the horizontal is called the *dip* or *inclination*. In the northern hemisphere, the north pole of the needle dips downward, or the dip is positive, while in the southern hemisphere, the south pole of the needle dips downward, or the dip is negative. The line of no dip, which encircles the earth near the equator, is called *the magnetic equator*. When the declination and the dip are known, the direction of the magnetic field is evidently determined.

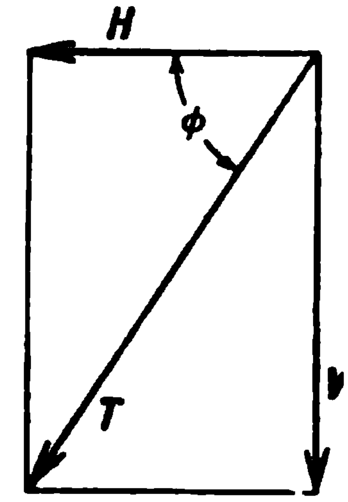


FIG. 250.

387. Intensity of the Earth's Magnetic Field. —The method of §383, for the determination of the intensity of a magnetic field by using the same magnet in deflection and oscillation experiments, applies to the earth's field. Evidently the intensity thus determined is that in the horizontal direction, or the *horizontal component* H of the earth's field. If we know the dip ϕ, and the horizontal component H, we get directly the total intensity, T, that is (Fig. 250),

$$T = H / \cos \phi$$

The vertical component V is also given by the relation

$$V / H = \tan \phi$$

388. Magnetic Maps.—The results of the magnetic surveys, in which the declination, the dip and the intensity of the earth's magnetic field at various places have been determined, are best shown by means of lines drawn on a map. A line drawn through points having the same declination is called an *isogonic line* Fig. 251 shows the isogonic and the agonic lines for the world. The amount of the declination is indicated by the figures on the line. Thus the line passing near New York is that of 10° W. declination. It is seen that the line, passing near Cincinnati, Ohio, has a declination of 0°. This is an *agonic line*.

A line connecting points having the same dip or inclination is called an *isoclinic line*. The isoclinics follow the general direction of the parallels of latitude, and some of them are indicated by dotted lines in Fig. 251.

Over a *magnetic pole*, the dip is 90°. A magnetic pole is not at the corresponding geographical pole. The one in the northern

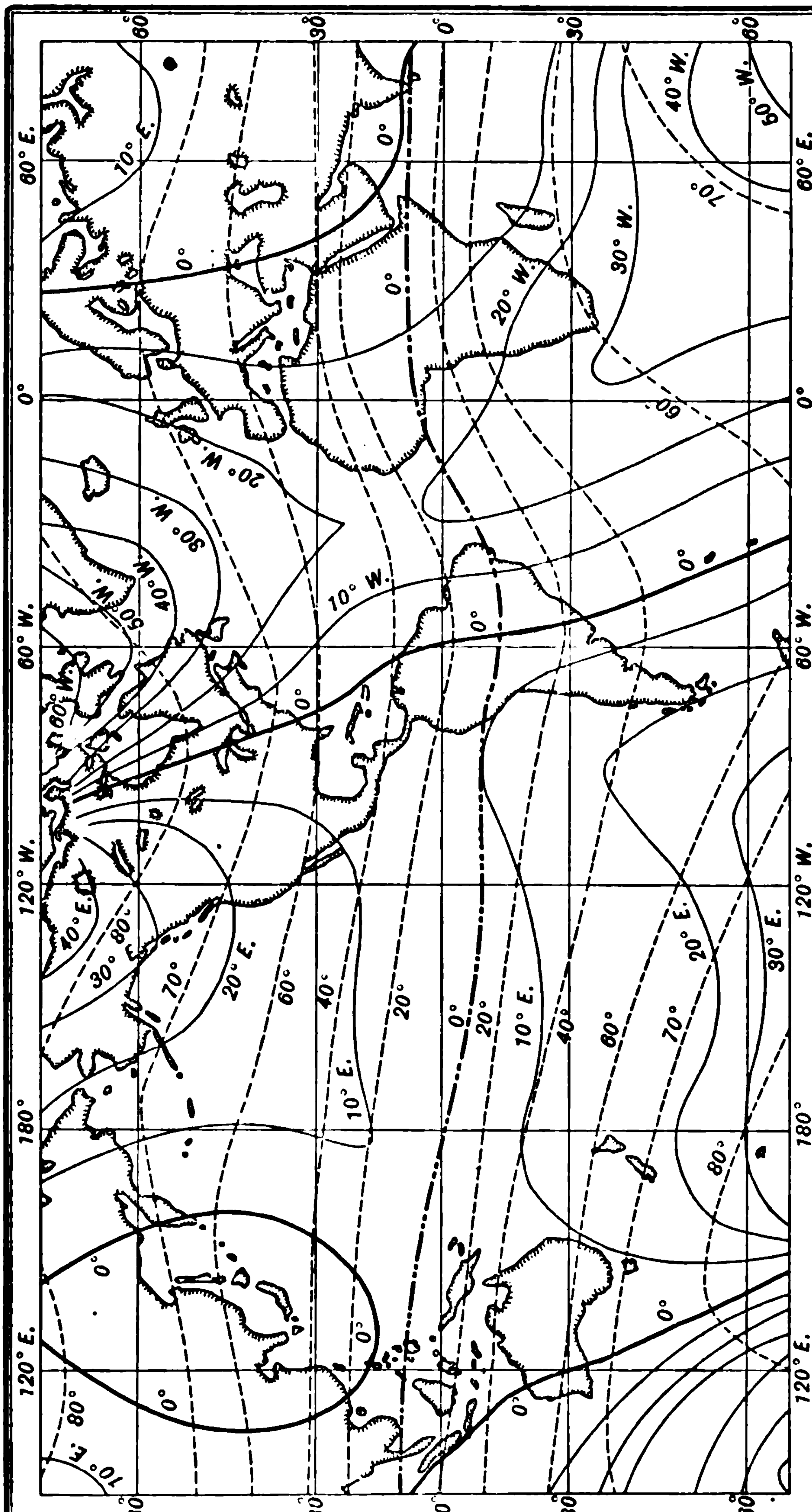


Fig. 251.
Magnetic map of the earth. Isogonic lines——— Isoclinic lines..........

hemisphere is at present in the neighborhood of 97° W. long., and 75° N. lat.; but the magnetic poles of the earth are not to be thought of as definite points.

A line connecting points of the same intensity is called an *isodynamic line.*

339. Time Variations of the Declination.—Observations of the magnetic declination have been taken in western Europe with more or less regularity since 1580. These observations show that in 1580 there was an easterly declination in London of 11° 15′ which decreased until it was zero in 1657, and then became

Fig. 252.

westerly, reaching a maximum westerly declination of about 24° 38′ in 1818; since that time it has been decreasing. In recent years the declination has been decreasing at about 5′ per year. This variation is called the *secular variation* of the declination. Observations also show similar secular variations of inclination and intensity. In addition to the above there are variations of the earth's magnetic elements, which have *annual* and *daily* periods. A very interesting fact in terrestrial magnetism is that times of disturbances on the surface of the sun are times of maximum changes in the earth's magnetism. Thus Fig. 252 given by Bigelow in the *U. S. Monthly Weather Review*, shows

that the eleven-year period of the sun spots corresponds to periods of maximum magnetic variations. The phenomena of the *aurora borealis* are also closely connected with magnetic disturbances.

Why the earth is a magnetized body has been a much debated question. Among the causes discussed have been, distributions of magnetized masses in the earth, and the presence of electric currents in the earth and in the atmosphere. No complete and satisfactory theory has, however, been reached. For more extended discussions of terrestrial magnetism, the student is referred to the article in the eleventh edition of the Encyclopædia Britannica and to various articles in the journal, *Terrestrial Magnetism.*

ELECTROSTATICS

390. Fundamental Experiments.—If a rod of hard rubber is rubbed with fur, it is found that light particles are attracted to the rod. Thus shreds of paper and pieces of pith cling to the rubber, and after a short contact are strongly repelled. A very convenient instrument for detecting this attraction and repulsion is a small gilded pith ball hung by a silk fiber. A body which has acquired this property of attracting and then repelling light particles is said to be electrified. The cause of this attraction is ascribed to an agent called "electricity," and the electrified body is said to have "a charge of electricity," or simply to be "charged." A suspended pith ball or other device for detecting electrification is called an electroscope.

This electrified state may be acquired similarly by other substances. Among the substances which show it very strongly are amber, rubber, resin, sulphur, sealing wax and shellac when rubbed with fur, and glass and crystals when rubbed with silk. It will be seen later that electrification results when any two different substances are rubbed together, but that, in most cases, it can be detected only by special appliances.

391. Two Kinds of Electrification.—If a rubber rod which has been electrified by friction with fur be suspended by a thread so that it is free to rotate in a horizontal plane about its middle point, it is found that this rod is repelled by a similarly electrified rubber rod. If next a glass rod be electrified by friction with silk,

and brought near the suspended electrified rubber rod, there is an attraction. In the same way, the electrified glass rod can be suspended, and it is found that it is repelled by another electrified glass rod. That is, the electrification of glass from friction with silk acts in an opposite way to the electrification of rubber from friction with fur; or, in other words, there are *two kinds of electricity*. The electricity on the glass is called *positive* electricity; while that on the rubber is called *negative* electricity. The above experiments show that *bodies charged with like kinds of electricity repel each other, and bodies charged with unlike kinds of electricity attract each other.*

We have seen that a rubber rod becomes electrified negatively by friction with fur. If we now test the fur, we find that it also is electrified, but that it attracts the electrified rubber, and repels the electrified glass rod. That is, the fur becomes electrified positively at the same time that the rubber becomes electrified negatively. In the same way, experiment shows that in the friction of glass and silk, the silk becomes electrified negatively, while the glass is being electrified positively. In general, when electrification is produced by the friction of two different substances, both substances are electrified, the one with one kind of electrification and the other with the opposite kind of electrification. In the following list, a number of common substances are arranged in a so-called "electric series," the order being chosen so that if a substance be rubbed with a second substance which is further down the series, the first substance becomes positively, and the second substance negatively electrified. Thus when glass is rubbed with silk, the glass is positive and the silk negative, while glass rubbed with fur becomes negatively electrified. The electrification of a substance, however, depends so largely upon the surface conditions, impurities, temperature, etc., that the order in the series is only approximate.

fur	glass	metals	resin
wool	silk	hard rubber	sulphur
quartz	wood	sealing wax	gun cotton

392. Transfer of Electricity, Conductors and Insulators.—If *all* points of an electrified rod be touched to a metal ball which is held in the hand, it is found that the rubber has lost all of its

electrification; and further, if the experiment is repeated, except that the ball is mounted on a glass or rubber stand, it is found that the ball has acquired the electrification of the rubber, and that it attracts and then repels light particles, such as the suspended pith ball. By touching the mounted ball to a second mounted ball, a portion of the electrification is again transferred. Electrification can thus be transferred from body to body by conduction. The repulsion of the pith ball after it has touched a charged body, is due to its becoming charged by conduction with the same kind of electricity as the charged body.

But the metal and the rubber differ in one very important respect. If the metal ball is touched at any one point by a wire or by the hand, and is thus connected to the earth through the wire or the human body, the whole ball loses its charge. But an electrified rubber or glass rod is not completely discharged unless every point of the rod is touched with the wire or with the hand. That is, electricity moves freely from point to point of the metal, but does not move readily along rubber or glass. This difference is further shown by joining the electrified mounted ball with a second mounted ball, first, by a glass or rubber rod, and second, by a metal rod supported by a rubber handle. The electrification is transmitted or conducted along the metallic connections but not along the glass or the rubber connection. Metals are thus seen to be good *conductors* of electrification or electricity, while glass and rubber are poor conductors or *insulators*. It is now apparent that the purpose of mounting a metal body on a glass or rubber support is to insulate it from the earth and other bodies.

Experiments show that no substance is a perfect insulator, and likewise that no substance is a perfect conductor of electricity. The best insulators are amber, rubber, sulphur, shellac, glass, porcelain, quartz, air, silk, etc.; the best conductors are the metals, acids, moist earth, etc. Dry wood, paper, cotton and linen thread, etc., are semi-conductors.

Gilbert called such substances as amber, rubber, glass, sealing wax, quartz, etc., *electrics*, after the Greek word for amber (ἤλεκτρον). It was known to the ancient Greeks that amber, when rubbed, acquired the striking property of attracting pith, straw, and other light bodies, but up to 1600 this was an isolated fact regarded as peculiar to amber and jet.

Gilbert showed that many other substances acted like amber when rubbed, and hence he called such substances, *electrics*, or amber-like bodies. Gilbert failed to find the same property in metals, when he rubbed them, and hence he called them *non-electrics*. It was not until many years later (1736) that Stephen Gray, another Englishman, showed that some substances were good conductors of electricity and other substances bad conductors or insulators. It was then possible to show that a metal body is readily electrified by friction, provided the metal is supported on an insulating stand. After this discovery the terms "electrics" and "non-electrics" lost their meaning, and in the present literature they have only a historical interest.

393. Electrification by Electrostatic Induction.—If an insulated conductor, *A*, be electrified, say positively, and brought near *B*, a second insulated conductor, *B* becomes electrified. This is shown by the repulsion of the small pith ball electroscopes which are attached to each end of *B*. By bringing a suspended gilded pith ball in contact with *A* and thus charging the ball with positive electricity, we can test the charges on *B*. It is found that the near end of *B* attracts, and the far end repels the pith ball; that is, the electrification on *B* is of two kinds, the far end having the same kind as that on *A*, and the near end having the opposite kind to that on *B*. If *A* is now moved to a distance from *B*, *B* is no longer charged, but becomes charged again when *A* is brought back. If *B* is now joined to the earth by a wire or by the hand, the charge at the far end of *B* disappears, but the charge on the near end remains. The charge on the near end is called a "bound charge," while that on the far end is called a "free charge." The "free charge" is one that escapes when joined by a conductor to the earth, while the "bound charge" does not so escape, because it is attracted by a charge of the opposite kind. If the connection of *B* with the earth is broken, and *A* removed, it is found that the charge on *B* distributes itself over the conductor and is "free." The conductor *B* is then electrified oppositely to *A*, while the charge on *A* is not diminished. The above process is called charging a body by *electrostatic induction* or influence.

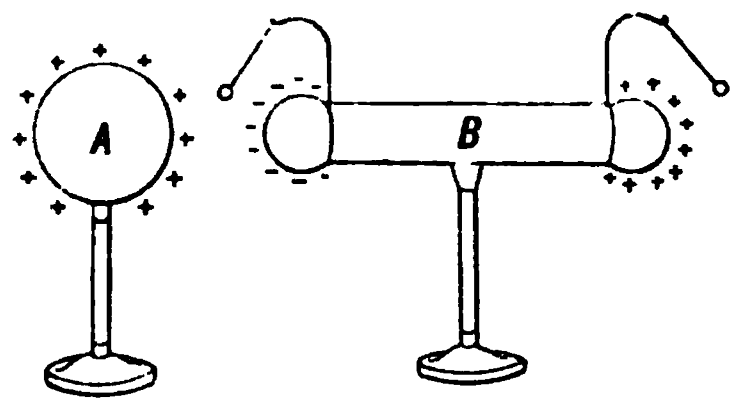


Fig. 253.

394. Theories of Electricity.—It has been stated that electrification is assumed to be due to an agent called "electricity." Various theories have been held as to the nature of electricity.

One of the first theories was that due to Benjamin Franklin and, although stated a hundred and sixty years ago, it is still, in the essentials, one of the most consistent theories of electricity. Franklin assumed that there is an "electrical matter," probably consisting of very fine particles, so light as to be practically imponderable or without weight, and that this electrical matter flows most freely, that is, it is a "fluid." This electrical fluid is distributed throughout all bodies, and each body has naturally a certain normal amount of it. If more than this normal amount is added to a body, the body is positively electrified; if the body by any means has less than its normal amount of the fluid, the body is negatively electrified. Further, "electrical matter differs from common matter in that the parts of electrical matter naturally repel each other," but they attract ordinary matter. Thus the process of electrifying rubber by friction is one in which the fur gets more than its normal amount of the electrical fluid, and the rubber less than its normal amount, while, in rubbing glass with silk, the glass gains electrical fluid at the expense of the silk. To electrify a body positively is thus simply to transfer from a second body a portion of its electrical fluid, and the second body will then have a deficit or will be negatively electrified.

Another fluid theory of electricity, that has been widely held, is Symner's two-fluid theory of electricity. Symner assumed that there are two electrical fluids, a positive fluid and a negative fluid. In its neutral or unelectrified condition, a body has equal quantities of these two fluids; when a body is electrified positively, it has more positive than negative fluid; and when electrified negatively, it has more negative than positive fluid. It is further assumed that the two fluids attract each other. It is evident that the above fluid theories are equivalent to each other, if we simply suppose Symner's negative fluid is the "deficit of the positive fluid." The Franklin theory has the advantage of assuming only a single fluid, and is more nearly in accord with the present electron theory of electricity.

The electron theory of electricity is Franklin's one-fluid theory

extended and made much more precise so as to account for numerous phenomena recently discovered. According to this theory electrification is due to negatively charged particles, called electrons or corpuscles, which are all precisely similar but very much smaller than the smallest atoms. In its natural unelectrified condition a body has a certain number of electrons; when it has more than this normal number, the body is negatively electrified, and, when it has less than the normal number, it is positively electrified. Different lines of research have shown that the mass of an electron must be about 1/1800 of the mass of a hydrogen atom. It seems probable that in a non-conductor most of the electrons are associated with or bound to atoms and possibly vibrate or rotate about the centers of atoms, as planets rotate about the sun; but in conductors most of the electrons are dissociated from atoms and are capable of moving about freely, thus accounting for the flow of electricity in conductors. While the body of evidence for the electron theory in some form is very great, the mechanism of the attraction between electrons and atoms which have less than the normal number of electrons remains as yet unexplained, and, to allow for this difficulty, it is still customary to speak of an atom as having a charge or "nucleus" of positive electricity which it cannot lose.

The fluid theory of electricity has in some form been used so long as a working hypothesis, that the terms of electrical science are based on the concept of a fluid. But in using such words as "flow," "current," etc., we do not commit ourselves to any particular theory.

395. Gold-leaf Electroscope.—The most sensitive and generally useful means of detecting electrification is the gold-leaf electroscope. In its usual form, it consists of two pieces of gold leaf hung beside each other from the lower end of an insulated metal rod. The upper end of the rod terminates in a ball or a plate. The gold leaves are enclosed in a case made wholly or partly of glass, for protection from air currents and so that the movement of the leaves can be observed. When the case is largely of glass, strips of tin foil are often pasted on the glass and connected through the base to earth for "screening" (§396).

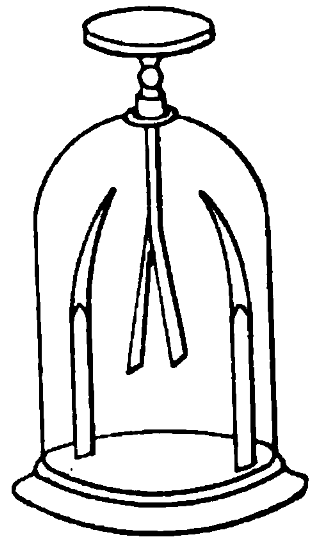


Fig. 257.

If the plate of the electroscope is electrified by contact with a charged body, the leaves, being charged with like kinds of electricity, diverge, and stay apart until the electroscope is discharged by connection with the earth.

The more usual method of charging the electroscope is by electrostatic induction. When a body which is charged positively is brought near the plate, the latter becomes charged with a "bound" negative charge and the leaves with a "free" positive charge. The free charge escapes when connection is made to earth, and the leaves collapse. The earth connection is now broken and the electrified body is then removed, thus freeing the "bound" negative charge. This spreads over the electroscope and the leaves diverge. The electroscope is thus charged negatively, that is, oppositely to the inducing charge. If now a positive charge is again brought up, the leaves collapse but, if a negative charge is brought up, the leaves diverge still further. Hence, if we know the kind of charge on an electroscope, we can determine the kind of charge on a body. If the leaves *first* converge as the body is brought up, then the body is charged opposite to the electroscope; if the leaves diverge as the body is brought up, then the body is charged with the same kind of electricity as the electroscope.

In a modified form of the gold-leaf electroscope (Fig. 255), a single strip of gold leaf hangs along a brass plate. The exact divergence of the gold leaf from the plate is easier to determine than the amount of divergence of two leaves, and so this form is better adapted for making measurements. The figure also shows devices to secure the highest insulation. The brass plate *P* with its gold-leaf strip *L* is supported separately by a sulphur bead *S*, and connection for charging is made by a special charging wire. The latter is a wire bent with two right angles, and fixed so that, by turning it, connection between the gold leaf and the upper disk can be made or broken.

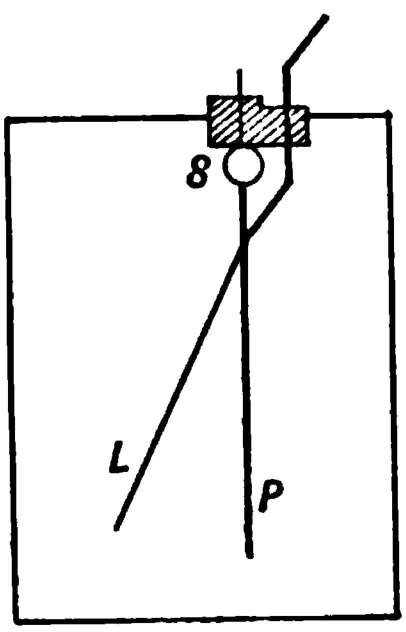


FIG. 255.

396. Electricity Confined to Surface of Conductors.—A very important law in the distribution of electrification on conductors is that it is all on the surface of the conductor. One method of showing this is by means of a hollow conductor in which there is a small opening. The

conductor is insulated and charged. If a small metal plate with an insulating handle, called a "proof plane," be now brought in contact with the various parts of the surface of the conductor, and then tested by bringing it to the gold-leaf electroscope, it is found to be charged. But if it be touched on the inside and brought to the electroscope there is no charge. That is, there is no charge on the inside of a conductor, unless the charge is insulated from the conductor.

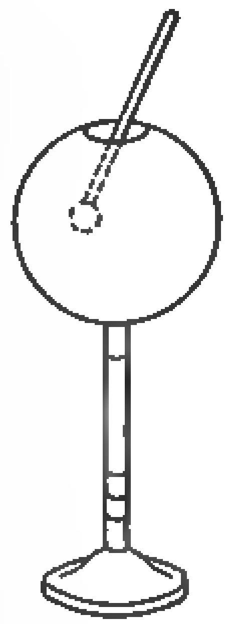


Fig. 256.

Another experiment showing this, is to charge an insulated metal body, and, after carefully introducing it through the opening of the hollow conductor, to touch it to the inside of the hollow conductor; on removing and testing the body, it is found to be completely discharged, its charge being now found on the outside of the conductor.

Another experiment is shown in Fig. 257. *A*, a metal sphere on an insulating stand, is charged. Two insulated hemispherical metal cups, *B* and *C*, are arranged so as to completely enclose and touch *A*. When *B* and *C* are removed, it is found that *A* is free from any charge and all the charge is on *B* and *C*.

Still another experiment showing the same law, is to put a sensitive electroscope inside a finely woven wire cage, connecting it with the cage. The insulated cage can now be strongly electrified and still the electroscope will show no charge on the inside. Faraday constructed a large metallic covered box which

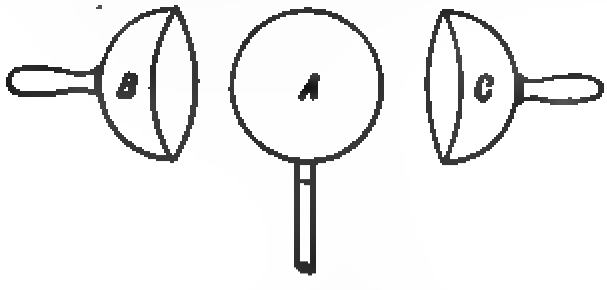


Fig. 257.

Fig. 258.

he insulated, and into it he carried his most sensitive electroscopes. He found that these showed no effects, even when spark discharges took place from the outside. Experiments with the thinnest of films show that the electrification is always on the surface.

The above experiments also show that a body can be shielded from outside electrical disturbances by surrounding it with a metal case. This is done frequently with measuring instruments, especially with electroscopes and electrometers. The explanation of the above facts in terms of lines of force will be given later (§399).

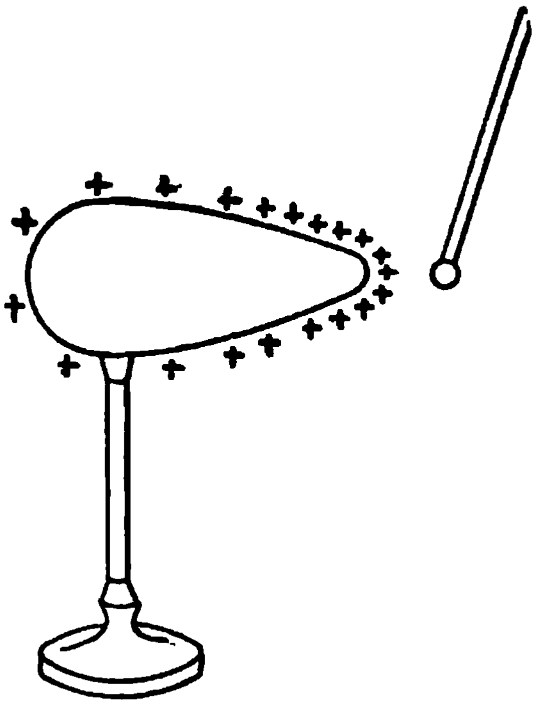


FIG. 259.

397. Distribution of Electrification on Conductors. *Effect of Points.*—We can investigate the distribution of electrification on the different parts of a conductor by means of a "proof plane" and a gold-leaf electroscope. For this purpose take an egg-shaped conductor and charge it. Touch the proof plane to various parts of the conductor and then test the proof plane by the electroscope (Fig. 259). It is found that the deflection of the electroscope is greatest when the "proof plane" has been in contact with the pointed end of the conductor, and least when it has been in contact with the flat parts of the conductor; or, in general, that *the electrification is greatest at parts of greatest curvature.*

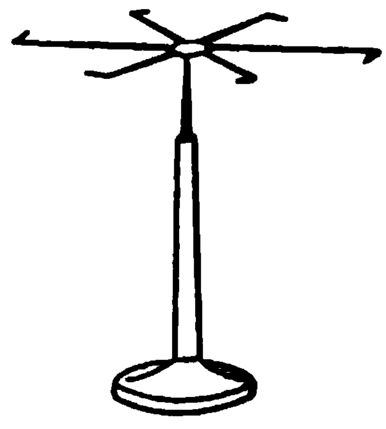


FIG. 260.

The curvature of a sharp point approaches infinity, and it follows from the above, that the electrification on such a point should become very great. That this is so is shown by the fact that an insulated conductor supplied with needle points discharges itself almost immediately. Also, if a needle point be held toward a charged conductor, the conductor loses its charge almost immediately. The induced electrification on the point is so great that it somehow breaks across and discharges the conductor. An interesting device for showing the discharge from points is "the electric wheel" (Fig. 260). It consists of a series of pointed wires inserted horizontally into a metal ball, which is balanced on a steel point, the pointed wires being arranged as shown, in

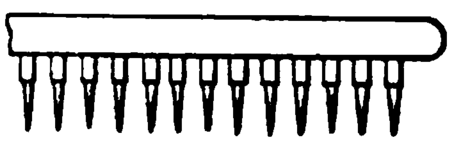


FIG. 261.

a "whirl." The discharge from the points causes a reaction which drives the wheel around rapidly. A metal rod carrying a row of metallic points (Fig. 261), and called a "comb," is used in static electrical machines to collect the charges from the revolving glass plates across a short air space (see §409).

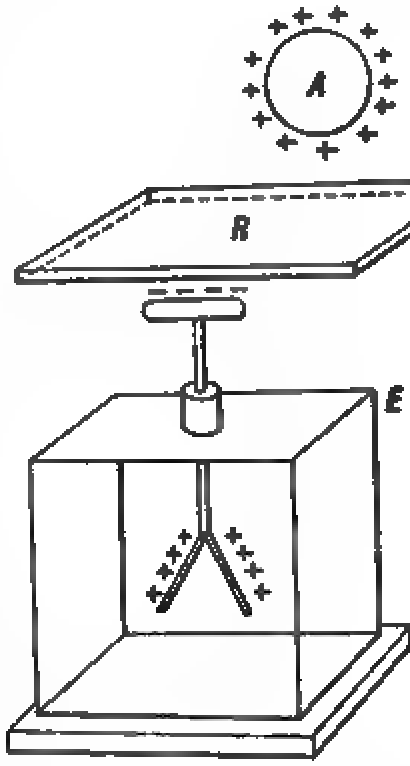


FIG. 262.

398. Fields and Lines of Electric Force.—The experiments which we have described above can all be explained, if we assume that one electric charge acts on another "at a distance," that is, directly across space without the action of an intervening medium. Thus, to explain electrostatic induction, we might say that the positive charge on a body *A* repels the positive charge in the gold-leaf electroscope to the leaves, and attracts the negative electricity to the plate, in this account making no mention of any intervening medium. But a simple experiment shows that the intervening medium cannot be neglected. If we introduce a thick plate of hard rubber, *R*, between the inducing charge on *A*, (Fig. 262), and the electroscope *E*, we see that the leaves come nearer together, and when the hard rubber plate is removed, they go back to their original position. The action is similar with plates of sulphur, shellac, glass and other insulators. That is, we see that the inductive action of *A* upon the electroscope, depends upon the intervening medium. Similar phenomena led Michael Faraday to a study of the insulating medium. (See §413.)

FIG. 263.

Faraday could not conceive that one body could act upon another otherwise than by a push or a pull through an intervening body or medium of some kind. Like Newton in the case

of gravitation he could not think of "a body acting where it was not." Faraday, therefore, called an insulating medium a *dielectric*, a word suggesting electric action through or across the medium, because he thought of the electric action as due to stresses in the intervening insulators. In accordance with this idea, he called the space about an electrified body *a field of electric force*, or *an electric field*. We can think of an electric field as a region in

Fig. 264a

which there are electric stresses, and these stresses can be indicated by lines or tubes. A line of electric force is the path along which a small positive charge tends to move. Thus, about an electrified sphere there are radial lines or tubes, indicating the direction of the stress in the field. We can map out these lines of stress by taking a small positively electrified body, and noting the direction of the force on the body at each point. In the case of a charged sphere hung in air at a very great distance from all other conductors, these lines apparently disappear where the forces become too small to detect, but if there are two conductors, (Fig. 264*a*), one charged positively and the other negatively, many of the lines that start from the positive charge will be

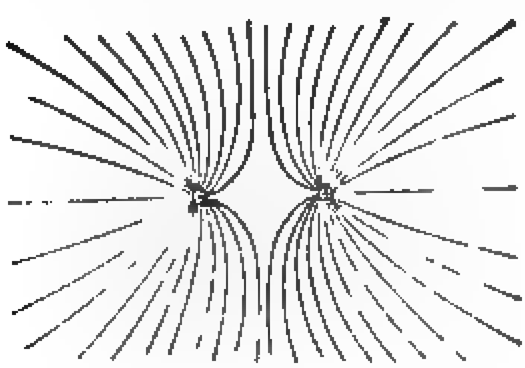


Fig. 264b.

found to terminate in the negative charge. Figure 264*b* shows the lines of force of the field due to two equal positive charges.

The mechanism of an electric field is not yet well understood Faraday regarded a line or tube of electric force as a chain of "polarized" particles of the intervening medium. By a "polarized" body is meant a body which has equal but opposite properties at its two ends or sides. Thus a bar magnet with its equal north and south poles, or a metal sphere with equal positive and negative induced charges is polarized.

According to the electron theory each polarized particle in a dielectric consists of an atom and its associated electrons. Being oppositely charged, they will tend to move in opposite directions in an electric field, but the separation will be very slight and will be limited by the attractions between them. The separation will be greater the greater the intensity of the field. We can thus form an instructive mental picture of what takes place in a dielectric, but this will not apply to an electric field in a vacuum and for this we have at present no explanation.

399. Faraday's Ice-pail Experiment.—Lines of electric force start from positive charges of electricity and terminate in negative charges and there is a negative charge somewhere corresponding to every positive charge. An experiment due to Faraday (Fig. 265) proves this to be the case. It is known as "the ice-pail experiment," because, when it was first performed, a metal ice-pail was used as the most convenient vessel at hand.

A metal vessel *B* with a narrowed opening is insulated and connected by a wire with the uncharged gold-leaf electroscope *E*. The ball *A*, hung by a silk thread and charged positively, is let down into *B*, but does not touch *B;* the electroscope becomes charged positively, and on the insde of *B* we have an induced negative charge. If *A* is removedi without touching *B*, the electroscope leaves contract, showing that the two induced charges unite and exactly neutralize each other. Now if *A* is again put inside, the leaves of the electroscope again diverge; and if now the ball is touched to *B*, *there is no change in the divergence of the electroscope.* When *A* is taken out, it is found to be discharged. Evidently the positive charge on *A* exactly neutralized the induced negative charge on the inside of *B*, and

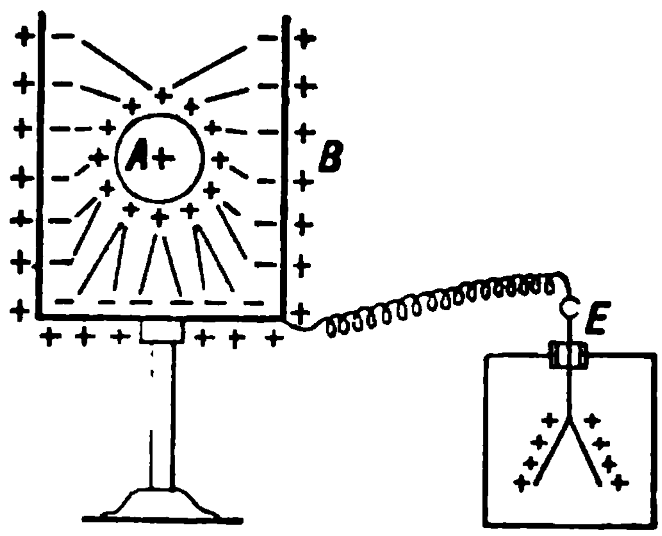


FIG. 265.

thus B and E remain charged positively. Hence the induced charge is equal and opposite to the inducing charge.

Since for every charge there is an equal opposite charge somewhere on the surrounding conductors (walls of room, earth, etc.), we must think of lines of electric force as always connecting these opposite charges, and also as being in a state of tension, so that they tend to contract and draw the two opposite charges together. The lines of force also seem to repel each other, as appears from the figures. This shows a lateral pressure in the medium which is very important in the theory of dielectric action.

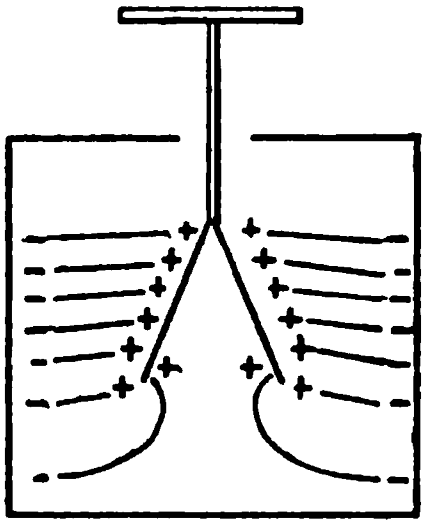


FIG. 266.

Electric repulsion is by this means resolved into two attractions in opposite directions. The repulsion between the two gold leaves in the electroscope is due to the tension of the lines of force between the charges on the leaves and the induced charges on the walls of the case (Fig. 266). The sensitiveness of the gold-leaf electroscope is thus changed by the presence of these neighboring conducting walls. The attraction of an electrified body for an uncharged conductor can now be seen to be due to the tension of lines of electric force. When the neutral body B is brought into the electric field of a positively charged body A, there is induced in B a negative charge on the near side and a positive charge on the far side, that is, lines of force connect A and B, as indicated in Fig. 267, and it is the resultant of the pulls of all these lines that causes the attraction.

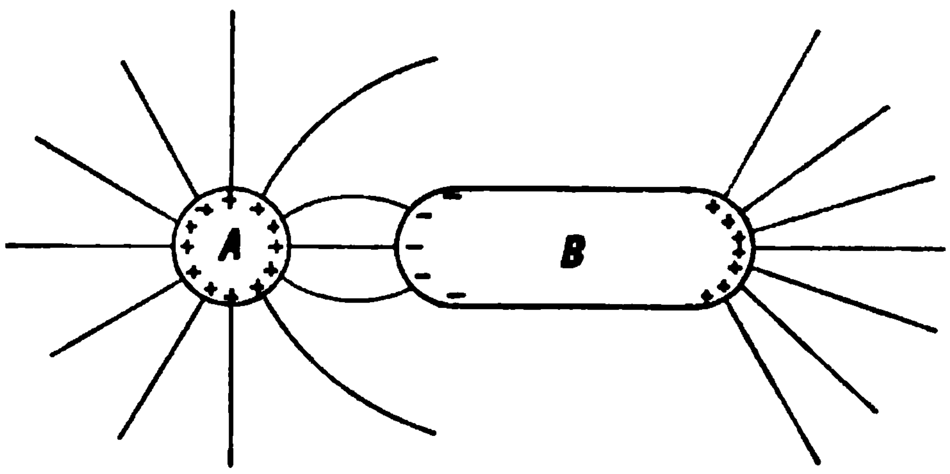


FIG. 267.

We have already seen that, in electrification by friction, the fur is electrified positively at the same time that the rubber is electrified negatively. By the "ice-pail" apparatus it can be shown that equal quantities of the two kinds are produced in the case

of friction. Fasten a small piece of fur, F, on an insulating handle, and rub the fur with a rubber rod, R, (Fig. 268). If both are inside the "ice-pail," the gold leaves indicate no charge; but when either is taken out there is a deflection. Hence the electrification on the fur is equal and opposite to that on the rubber.

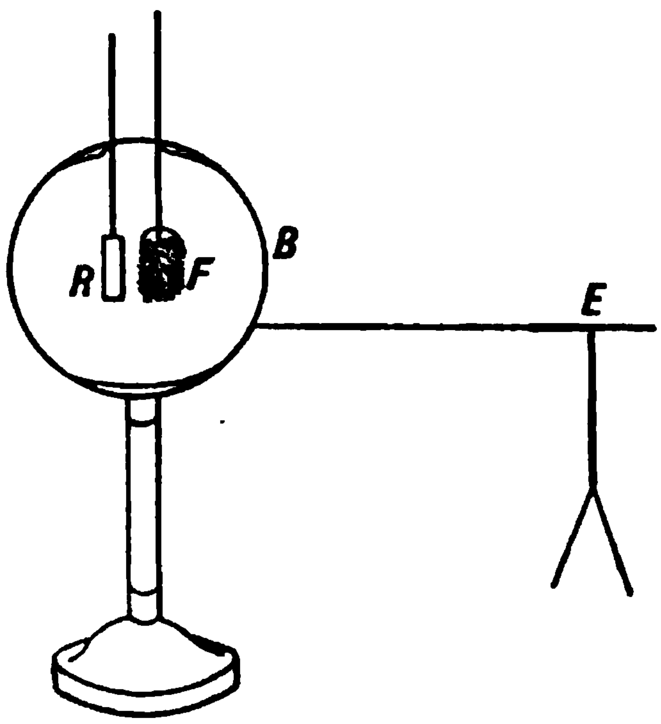


FIG. 268.

The experiments with the ice-pail apparatus show that in the charging and discharging of bodies there is no creation or destruction of electricity, but simply transfers of electricity. The total quantity of electricity remains unchanged. This fact is known as the *conservation of electricity*, and is in accord with the fluid theories, and also with the modern electron theory.

400. Energy of Charged Bodies.—The energy which an electric charge represents comes from the work done in separating the two kinds of electricities, and is equivalent to it. Thus, when we rub a rod of rubber with fur, the electrification is in no way proportional to the friction. A very light but complete contact is in fact all that is needed and a perfect contact is the only aim in rubbing the bodies together. But, to separate the fur with its positive charge from the negatively charged rod, work must be done against the mutual attraction of the two charges, or, using Faraday's concept of lines of force, the work is done in setting up stresses in the intervening medium, these stresses being represented by the lines or tubes of force stretching between the fur and the rod (Fig. 269). When the two charges come together again, the lines or tubes contract, and do work. The energy on this view is analogous to that of stretched elastic bands connecting two bodies.

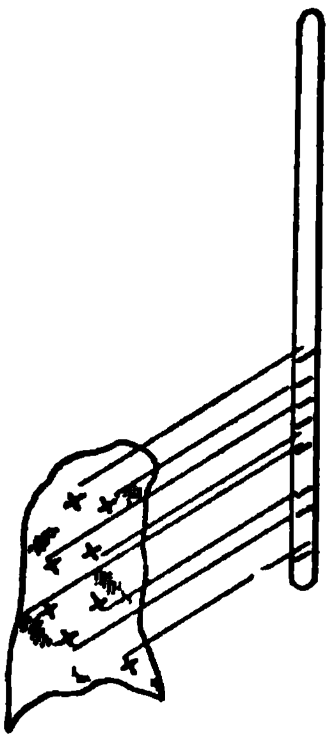


FIG. 269.

401. Law of Electrical Force.—The law which states how the force between two electrical charges depends upon the charges

and upon the distances between them, was first published by Coulomb in 1785 and hence is known as Coulomb's law of electric force. The law states that *the force between two electrical charges* varies (*a*) *inversely as the squares of their distance apart, and* (*b*) *directly as the product of the two electrical charges.* This is expressed by the formula,

$$F = \frac{1}{K}\frac{qq'}{r^2}$$

where F is the force, q and q' the two charges, r their distance apart, and $1/K$ a constant depending upon the units used and also upon the intervening medium. If we use the dyne as the unit of force, the centimeter as the unit of length, and define the unit charge or unit of electric quantity q, as follows: *Unit electric quantity is that quantity which at one centimeter distance in air exerts a force of one dyne on an equal quantity;* then the formula for charges in air becomes:

$$F = qq'/r^2$$

In the case of air, K has, by the definition of unit electric quantity, the value unity. For other intervening media or dielectrics, the more general formula must be used. The values of K for several dielectrics are as follows: (compare §413)

Air	1.00
Petroleum oil	2.07
Turpentine	2.23
Distilled water	75. +

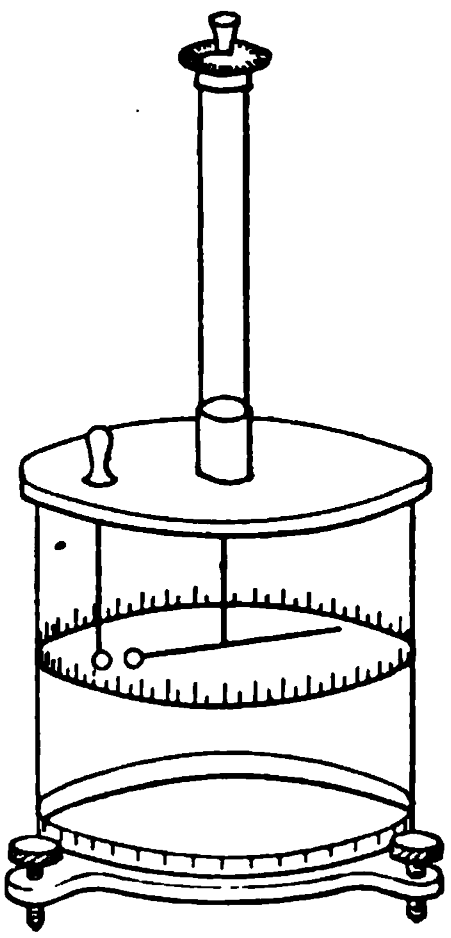


FIG. 270.

Coulomb arrived at the above law by experiments with his torsion balance (Fig. 270) similar to those by which he discovered the law of magnetic forces (§372), but the best proof of the law is the indirect one, that it is the only relation that explains exactly electrostatic phenomena. In fact, Henry Cavendish before 1785, thus established the law for his own use, though his papers were first published nearly a century later, long after others had reached the same results. This proof is based on the experimental fact that at any point, O, inside a hollow con-

ducting sphere there is no electric force from charges on the surface of the sphere. Draw straight lines through O, dividing the whole sphere into pairs of cones of small angular opening, and with a common apex at O, and with bases S_1 and S_2, etc., cut out of the spherical surface, the heights of these cones being r_1, r_2, etc. Let Q_1 and Q_2 be the charges on the bases S_1 and S_2. Since the electricity is distributed uniformly on the sphere, the charges Q_1 and Q_2 are proportional to the areas S_1 and S_2, or $\frac{Q_1}{Q_2} = \frac{S_1}{S_2}$. But the cones have the same angle and hence the bases S_1 and S_2 are to each other as the squares of the heights, or $\frac{S_1}{S_2} = \frac{r_1^2}{r_2^2}$, or $\frac{Q_1}{r_1^2} = \frac{Q_2}{r_2^2}$. Thus, the forces being equal and opposite, the resulting force at O on a test unit must be zero, if the inverse square law holds. Since these equations hold for all values of r_1 and r_2, it is evident that the inverse square law is the only one meeting the conditions of this experiment. Maxwell and others have repeated this experiment with very sensitive electroscopes so that the law of the force varying inversely as the square of the distance is one of the best established laws of electrostatics.

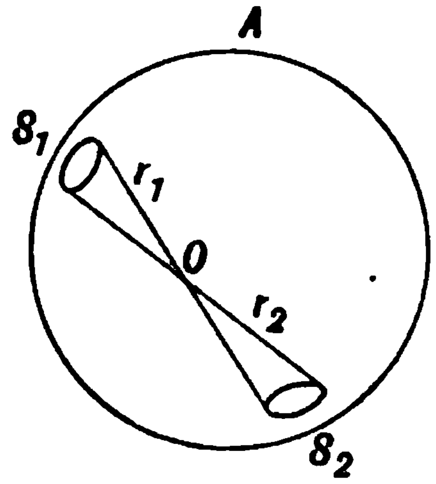


FIG. 271.

402. Electrical Potential.—To describe and explain the movement of electricity the terms "electrical potential" and "differ ence of electrical potential" are used. Thus if we join two conductors A and B by a wire and find that there is a flow of electricity from A to B, we ascribe this to a "difference of electrical potential" between the two bodies, and say that there is a flow from A to B because A is at "the higher potential" and B at "the lower potential." In the case of electrostatic charges, such as we have been describing, this flow or current is only momentary, because the two bodies come in an instant to the same "potential." By means of batteries and dynamos, as we shall see later, it is possible to maintain a continuous "difference of potential" and hence a continuous electric flow or current between two points of a conductor. The movement of a charged body from one point to another point in an electric field is also described as due to a "difference of electrical potential between

the two points" of the field. The positively charged body tends to move from A to B, because the point A is at "a higher electrical potential" than the point B.

Potential as used above is very analogous to *pressure* in fluids. Thus flow of a gas takes place from a tank of higher pressure to a tank of lower pressure when the tanks are connected, and the flow continues until the pressures are equalized. Another very useful analogy is that of *level* in liquids. A liquid tends to flow from points of higher level to points of lower level; to maintain the flow, the difference in level must be maintained. When a liquid flows from a higher level to a lower level, it loses some of its potential energy by transformation into energy of some other form. In fact the potential energy of a system always tends to a minimum. (§107.)

Let us apply this to the case of a charge in an electric field. Consider the electric field (Fig. 263, §398), about a positively charged sphere in air all other bodies being supposed to be at indefinitely great distances. A positive unit charge at a point x in this electric field has a certain potential energy, V_x, this being the number of ergs of work that the charge can do by moving under the action of the forces of the field from x to infinity (that is, completely out of the field). It is also equal to the work that is required to move the unit charge from an infinite distance up to the point x. The potential energy of a positive unit charge at a point x we call "the electrical potential at the point x," and represent it by V_x. In the same way, the electrical potential at a point y, or V_y, is the potential energy of unit charge at y. The potential energy of a charge in an electric field, as in other cases of potential energy (§63), evidently depends only on the final position of the charge, and not upon the path. That is, each point in an electric field has a definite electrical potential. Hence the difference of electrical potential of two points x and y, that is, $V_y - V_x$, is a definite quantity, and is defined as follows: *The difference in electrical potential between two points x and y in an electric field is equal to the number of ergs required to move unit positive charge from the point x to the point y.* If this work is positive, that is, if external work must be done on the positive charge to move it, then the potential of y is higher than the potential of x.

We thus see that it is in accordance with the principle of minimum potential energy that a positive charge tends to move from a point of higher electrical potential to a point of lower electrical potential. In the above we have assumed that the intensity of the electric field was not appreciably changed by the presence of the unit charge.

403. Zero Potential, Positive and Negative Potential.—In the case of water levels we choose some arbitrary level as a reference or zero level, the level of the sea being so chosen by universal agreement. All levels above sea-level are marked positive or plus (+), and all levels below sea-level are marked negative or minus (−). In an analogous way, the electrical potential of the earth is taken as the zero potential. Since the earth is a conductor, all points on it for electrical equilibrium are at the same potential; otherwise there would be an electrical flow until equilibrium was reached. A body A is thus at a positive electrical potential when positive electricity tends to flow from A to the earth; and in the same way, a body B is at a negative electrical potential when positive electricity tends to flow from the earth to the body B. To say a conductor has a "free" positive (or negative) charge is equivalent to saying that it is at a positive (or negative) potential.

The electrical potential, V, of a point or of a body with reference to the earth is therefore equal to the number of ergs required to move unit positive electricity from the earth to the point or to the body. If this work is positive, that is, if positive work is done *on* the test unit, then the potential is positive; but if the work is done *by* the test unit, then the potential is negative.

404. Equipotential Surfaces.—In an electrical field all points which have the same potential lie on an equipotential surface. To determine if two points are on an equipotential surface is to determine if work is done against electrical forces in the movement of a charge from the one point to the other, or whether a test charge tends to move from either point to the other. Lines of force always cut an equipotential surface at right angles; otherwise there would be a force component along the surface, which is not possible if a charge does not tend to move along the surface. Hence in the case of the electric field about a single charge distant from other charges, the equipotential sur-

faces are concentric spherical surfaces. In the figures of electric fields (§398), the equipotential surfaces are indicated by the dotted lines at right angles to the lines of force. The surface of a conductor is evidently an equipotential surface, if the electric charges on it are at rest; and hence lines of force enter and leave a conductor at right angles to the surface.

405. Units of Quantity and of Potential.—The unit of electric quantity has been defined (§401) as *the quantity which at one centimeter distance in air exerts a force of one dyne on an equal quantity.* This is the *c.g.s. electrostatic unit quantity.* For practical measurements a much larger unit called the "*coulomb*" is used. We can for the present define the coulomb as follows: 1 coulomb $= 3 \times 10^9$ c.g.s. electrostatic units of quantity.

The c.g.s. electrostatic unit difference of potential exists between two points when one erg of work is done in the movement of a c.g.s. electrostatic unit of charge from the one point to the other point. Hence to move q c.g.s. electrostatic units of electricity from a point of potential V_1 to one of potential V_2 takes $q(V_2 - V_1)$ ergs of work, or

$$W = q(V_2 - V_1).$$

In practical measurements of difference of potential the *volt* is used as the unit; 1 volt $= 1/300$ or $1/3 \times 10^{-2}$ c.g.s. electrostatic units difference of potential.

From the above, it follows that the movement of a coulomb of electricity against a *D. P.* of one volt, represents $3 \times 10^9 \times 1/3 \times 10^{-2}$ ergs $= 10^7$ ergs $= 1$ joule (§55). Hence

$$W(\text{joules}) = Q(\text{coulombs}) \times D.\,P.\,(\text{volts}).$$

In the special section on electrical units (§547) the reason for the choice of the above practical units will be discussed.

406. Potential Calculations.—The difference of potential between two points a and n due to a charge Q is given by the formula $V_a - V_n = Q\left(\frac{1}{r_a} - \frac{1}{r_n}\right)$, where r_a and r_n are the distances of a and n from Q and the medium is for the present assumed to be air. For let a and n and Q be in a straight line. (Fig. 272.)
The force on unit charge at a is Q/r^2_a, and the force at b, a point near a is Q/r^2_b. The average force between a and b can be taken

as $Q/r_a r_b$. The work done by the field in moving unit charge from a to b is then $(r_b - r_a)\, Q/r_a r_b$. Hence it follows that

$$V_a - V_b = (r_b - r_a) Q/r_a r_b = Q(1/r_a - 1/r_b).$$

Similarly for a series of neighboring points

$$\begin{aligned} V_a - V_b &= Q(1/r_a - 1/r_b) \\ V_b - V_c &= Q(1/r_b - 1/r_c) \\ &\dots\dots\dots\dots \\ V_m - V_n &= Q(1/r_m - 1/r_n). \end{aligned}$$

Summing these up, we get

$$V_a - V_n = Q(1/r_a - 1/r_n).$$

If the point n is at an infinite distance then $r_n = \infty$ and $1/r_n = 0$; and hence the potential of the point a is $V_a = Q/r_a$. Similarly the potential at any point x in the field of Q is $V_x = Q/r_x$, and the potential difference between a and x is $V_a - V_x = Q(1/r_a - 1/r_x)$

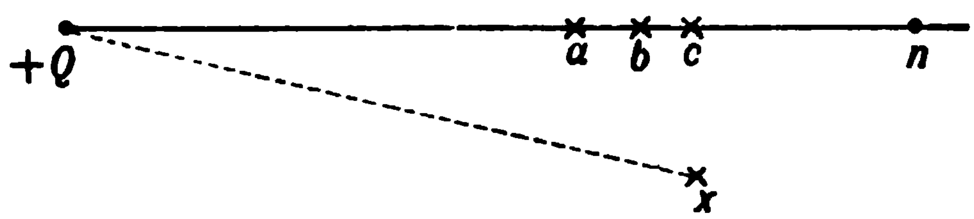


FIG. 272.

This result does not depend upon the path between a and x (§63). In the case of a number of charges Q', Q'', Q''', etc., the potential at a point a is the sum of the potentials due to each charge, that is

$$V = V' + V'' + V''' +, \text{ etc.}, = Q'/r'_a + Q''/r''_a + Q'''/r'''_a +, \text{ etc.}$$

When the medium is not air, the right-hand side of each of the above equations must be multiplied by the proper value of $1/K$ for that medium (§401).

The proof of the above by calculus is very simple. We have the differential of work $dW = Fdr = Q/r^2 dr$. Integrating between the limits r_a and r_n, we get the total work, or $V_a - V_n =$

$$\int_{r_a}^{r_n} dW = \int_{r_a}^{r_n} \frac{Q}{r^2}\, dr = Q\left(\frac{1}{r_a} - \frac{1}{r_n}\right)$$

407. Electrometers.—An instrument for measuring difference of electrical potential by means of electrostatic force, is called an electrometer. In these instruments there is a movable part—a charged needle or a disk—which is acted on by an electric field produced by the difference of potential to be measured. The common forms of electrometers are the quadrant, the disk or absolute electrometer, the single and the multi-cellular "electrostatic voltmeters." All of these were first developed by Lord Kelvin.

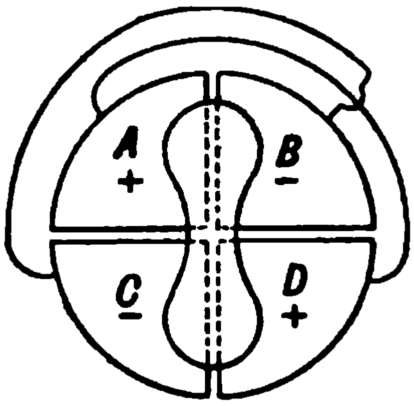


FIG. 273a.—Quadrants.

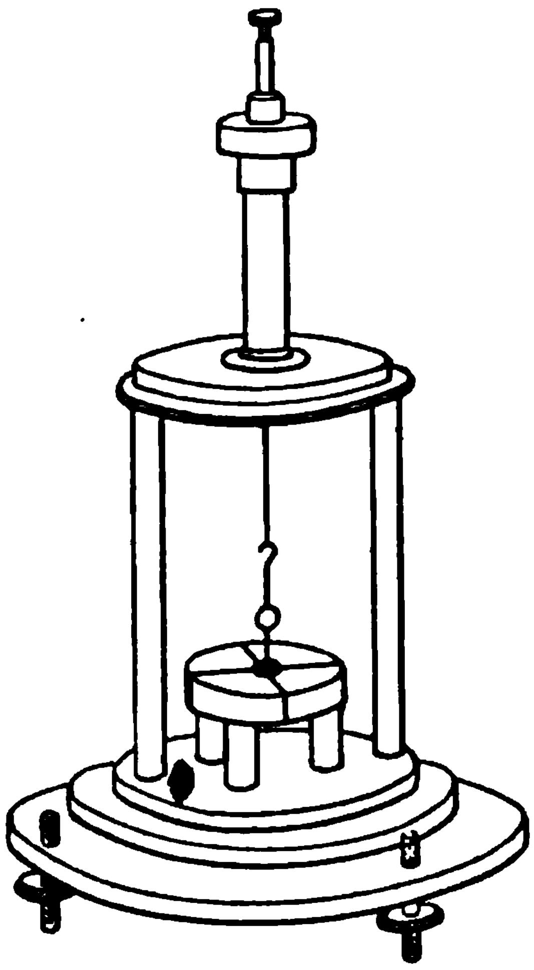


FIG. 273b.—Quadrant Electrometer.

The quadrant electrometer (Fig. 273*b*) consists of a light needle made of sheet aluminum or of silvered paper, which is suspended by a fine metal strip, or by a quartz fiber, inside of a shallow circular metal box. This metal box is divided into four quadrants which are mounted on insulating columns, preferably of amber. The diagonally opposite quadrants are connected by wires. The needle is free to move in a horizontal plane, and, in its position of equilibrium, the needle hangs along the line of separation of the two pairs of quadrants. Any deflection of the needle can be read by the movement of a beam of light reflected from a small mirror attached to the suspended needle. The needle is charged to a high positive potential, generally by joining it through its suspension to a high potential battery. So long as the two pairs of quadrants are at the same potential, there is no deflection of the needle. But, if the pair of quadrants A and D are at a higher potential than the quadrants B and C, there will be a couple deflecting the needle toward the quadrants at the lower potential. This couple is balanced by the torsion of the suspension, and it can be shown that for small angles, the difference of

potential is proportional to the angle of deflection. The quadrant electrometer is very sensitive, a common sensitiveness for the Dolezalek form (Fig. 273*b*) being a deflection of 1 mm. at 1 meter scale distance for a difference of potential of .002 volt.

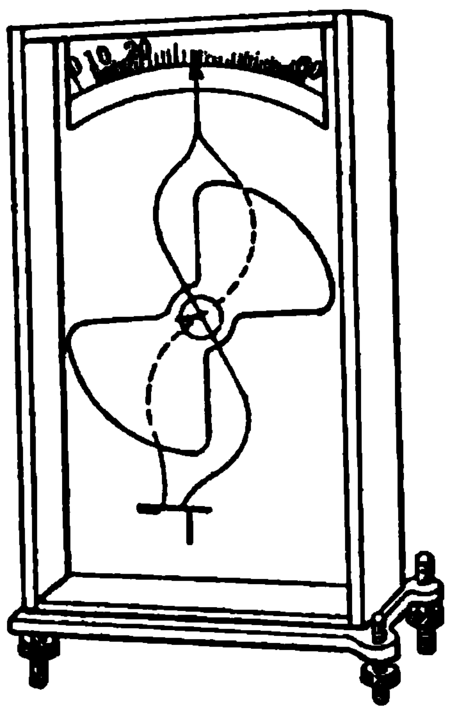


Fig. 274.—Electrostatic Voltmeter.

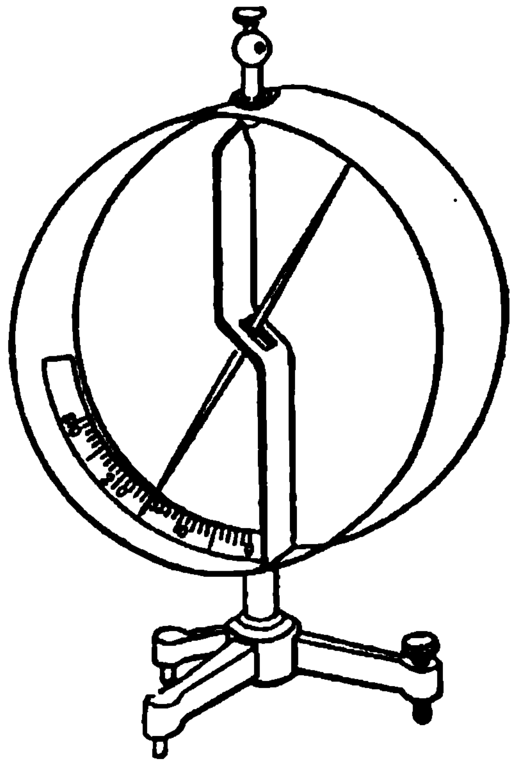


Fig. 275.—Braun Electrometer.

The quadrant electrometer can be used without charging the needle independently. The needle is then joined to one pair of quadrants, say *A* and *D*, so that the only charges are those due to the difference of potential to be measured. This arrangement is called the *idiostatic*, while the previous arrangement is called the *heterostatic*. The idiostatic arrangement is adapted for measuring larger differences of potential. In the "vertica-electrostatic voltmeter" (Fig. 274) there is a single pair of vertical quadl rants, and the needle is an aluminum vane balanced on knife edges, the dif-erence of potential between the quadrants and the needle is indicated by the tilting of the needle as shown by the pointer and scale. Differences of potential of from 1000 to 20,000 volts can be measured with this instrument.

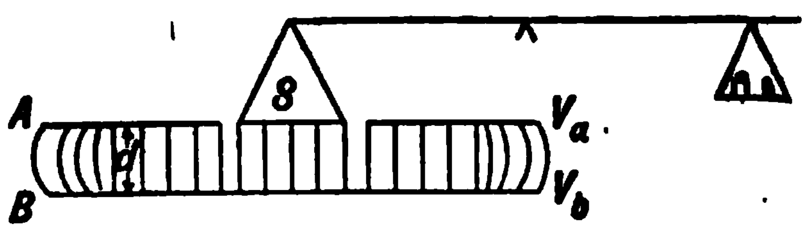


Fig. 276.

In the "multi-cellular electrostatic voltmeter," the needle consists of a series of parallel vanes, and wings horizontally between a corresponding series of fixed plates or quadrants. Increasing the number of vanes and quadrants, increases the ensitiveness, so that multi-cellular voltmeters reading as low as 10 volts are listed by the makers.

The Braun electrometer (Fig. 275) has a light needle that is pivoted on a horizontal axis, and the action is similar to that of the modified gold-leaf electroscope mentioned in §395.

In the disk electrometers there are two parallel plates, *A* and *B*, charged to the potentials V_a and V_b. Part of the upper plate, the disk

S, is movable and hung on a balance arm as in Fig. 276 (or it may hang by a calibrated spring), so that the force pulling it toward the plate B can be counterbalanced and thus measured. The outer part of the plate A is called the "guard ring," and serves the purpose of making the electric field uniform opposite the movable disk S, as indicated by the parallel lines of force. It can be shown that in this case, the difference of potential is given by the formula

$$\left(V_a - V_b\right)^2 = \frac{8\pi d^2 F}{S},$$

where d is the distance in centimeters between the plates, F is the force in dynes acting on S, and S the area in square centimeters of the disk. The difference of potential $(V_a - V_b)$ is accordingly given in absolute c.g.s. units and so this is an "*absolute electrometer.*"

It is now seen that the deflection of the gold-leaf electroscope is due to the difference of electrical potential between the gold-leaf and the metal case (Fig. 266, §399), since the lines of force connect the leaf and the case. The metal case is ordinarily connected with the earth and so is at zero potential. For small deflections, the difference of potential (V_1-V_0) is proportional to the deflections.

Static Electrical Machines

408. The Electrophorus.—The electrophorus is an instrument devised by Alexander Volta in 1790, to multiply electric charges

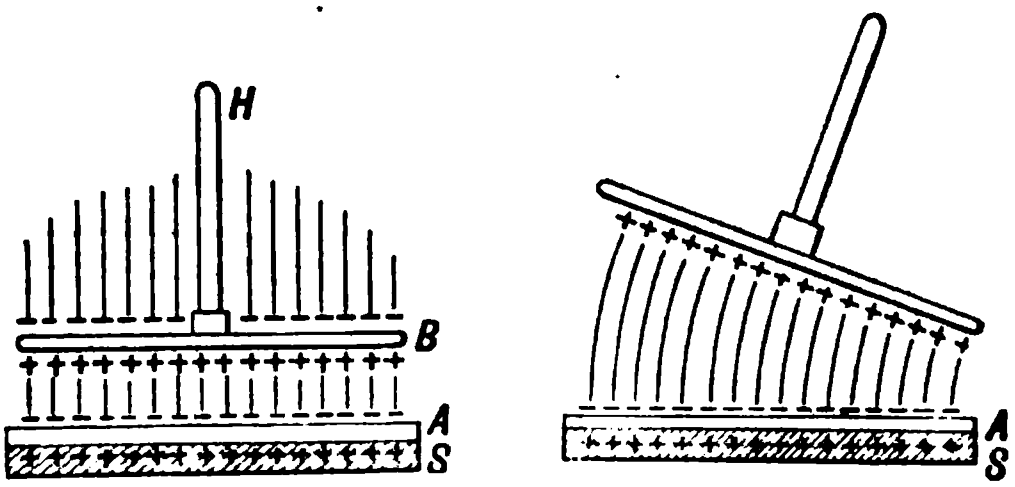


FIG. 277.

by electrostatic induction. It consists of a resin plate, A, on a metal plate or "sole," S, and a metal disk, B, with an insulating handle, H, at its center. The resin is charged negatively by friction. The metal base then has a "bound" positive charge which helps to hold the negative charge on the resin. Bring the metal disk B near and opposite the resin plate. There is induced in B a free negative charge and a bound positive charge.

The free charge is removed by connecting B with the earth for an instant. If the plate B is then moved away from the resin plate A, the positive charge is made a free charge and can be transferred to another conductor, such as an insulated sphere. This process can be repeated, the plate B being charged each time without decreasing the charge on the resin plate.

As described above, the plate B is brought "near" A, but in practice, the plate B is actually rested on A. The points of contact are, however, only comparatively few, so that little of the negative charge of the resin is removed to B, while the charge induced on B at the shorter distance is increased. Instead of a resin plate, a plate of sulphur, hard rubber, shellac, or other insulator can be used.

The positive charge on the plate B, represents energy. The source of this energy is in the work done in lifting the plate B after the free negative charge q has escaped. Work is done in pulling apart the charge of the plate and the disk, that is, in drawing out the lines of force connecting the plus and minus charges. The energy lies in this tension in the lines of force in the dielectric.

409. Electrostatic Induction Machines.—The earliest machines for producing electrical charges by rotation were friction machines. Thus a glass plate was electrified by rotating it between suitable rubbers, and the charge removed by metal brushes or pointed conductors. Such friction machines are seldom found now outside of museums, and have only historical interest. They have been superseded by the electrostatic induction machines. The two most common of these machines are the Toepler-Holtz, and the Wimshurst machines.

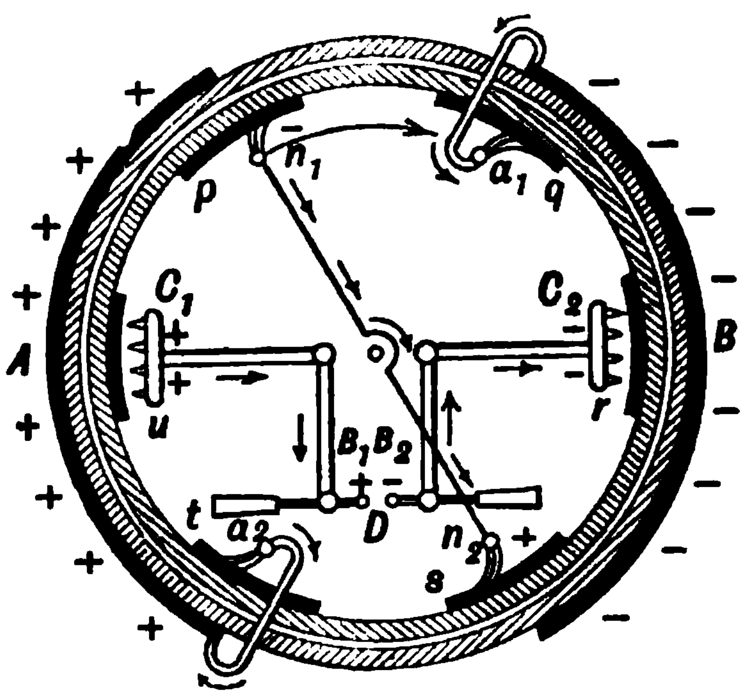


FIG. 278.

The Toepler-Holtz or Voss machine has two vertical glass or vulcanite disks, of which one is fixed and the other rotates about a horizontal axis normal to its center. The positions of

the various parts can be seen in Fig. 278 (due to Prof. S. P. Thompson), in which circles are used to represent the two disks; the farther disk which is fixed, is represented by the outer circle, and the nearer and rotating disk is representd by the inner circle. On the far side of the fixed disk are the combined paper and tin-foil "inductors," A and B. On the near face of the rotating disk, there are six tin-foil disks p,q,r, etc., called "carriers." These are spaced at equal distances around the disk. A "neutralizing rod," n_1n_2, reaches across the front face of the rotating disk, and by small brushes connects the two opposite carriers as they pass under the rod. On each inductor there is a metal brush arranged so as to make contact with each carrier as it passes. Collecting "combs" (§397), C_1, and C_2, with discharge rods and balls, D, are as shown, arranged in front of the rotating plate.

The action of the machine is as follows: One of the inductors, say A acquires by friction a small initial positive charge. This induces on the carrier at p, a bound negative charge and a free positive charge. The positive charge escapes along the neutralizing rod. The negative charge becomes free as the carrier passes to the position q. Here it shares its negative charge with the inductor B. At r it loses its negative charge to the comb C_2 and thus the ball B_2 is charged negatively. At s the carrier has a bound positive charge by induction from the negative inductor B, the free negative charge escaping along the neutralizing rod. At t the carrier shares its charge with the inductor A, and at u it loses its positive charge to the comb C_1. The ball B_1 is thus the positive terminal. The difference of potential between B_1 and B_2 is thus increased until a spark discharge takes place. A Leyden jar is usually connected to each terminal, the effect of which is

FIG. 279.

to increase the quantity of discharge for a given potential difference (see condensers §410). In a new machine, made by Wehrsen, the rotating plate of ebonite is triple and contains embedded metal sectors connected with the carriers, thus greatly increasing the output.

In the Wimshurst machine (Fig. 279) there are two parallel glass disks, geared to rotate in opposite directions. On each disk is a large number of tin-foil sectors, and each sector serves in turn as inductor and carrier. The neutralizing rods and combs are symmetrical on the two sides. The action of the machine is similar to that of the previous machine and can be followed from the + and − signs on the figures.

410. Electrical Capacity.—If two metal balls of different sizes be put in contact and charged, they will be at the same electrical potential, but they will not have the same electrical charges. This can be shown by hanging each ball separately in a metal cup on the plate of a gold-leaf electroscope and noting the divergence of the leaves (Fig. 265, §399). The fact that it takes more electricity to raise the potential of A a certain amount than to raise the potential of B the same amount, we describe by saying that the "electrical capacity" of A is greater than the electrical capacity of B.

Electrical capacity may be illustrated by the capacity of a tank for gas. The mass of gas (neglecting temperature changes) depends upon two things, (a) the dimensions of the tank, and (b) the pressure of the gas. The mass, M, of the gas equals the pressure, P, of the gas, multiplied by K, *the mass of the gas in the tank at unit pressure*, or $M=KP$. Similarly, the electrical charge Q on a conductor is equal to the electrical potential V, multiplied by C, *the electrical charge on the conductor at unit potential*, or $Q=CV$, assuming that V is due entirely to Q, that is, that the conductor is not in a field due to other charges. In this statement, C is called the *electrical capacity* of the conductor; *it is the quantity of electricity required to raise the potential of the conductor by unit amount.*

We can thus compare the electrical capacities of two conductors A and B, either, (1) by comparing the charges required to raise them to the same potential, or (2) by comparing the potentials to which equal charges raise the two conductors. In

the latter case, the greater the capacity, the less a given electrical charge would raise the potential.

It can be shown that the electrical capacity of a conductor depends not only (*a*) upon its size, but also (*b*) upon its shape, and (*c*) upon the position of neighboring conductors, and (*d*) upon the surrounding insulating medium or dielectric. That the capacity of a conductor depends upon the shape can be demonstrated by the apparatus shown in Fig. 280. The conductor is connected by a wire with an electroscope. A charge Q is given to it, and this raises the potential of the system to V, as indicated by the divergence of the electroscope. The conductor, being made of a series of cups, can now be drawn out, thus changing the shape of the conductor. The electroscope leaves converge, indicating a lowering of potential, although the charge on the conductor is not changed. Upon restoring the shape of the conductor the potential is restored.

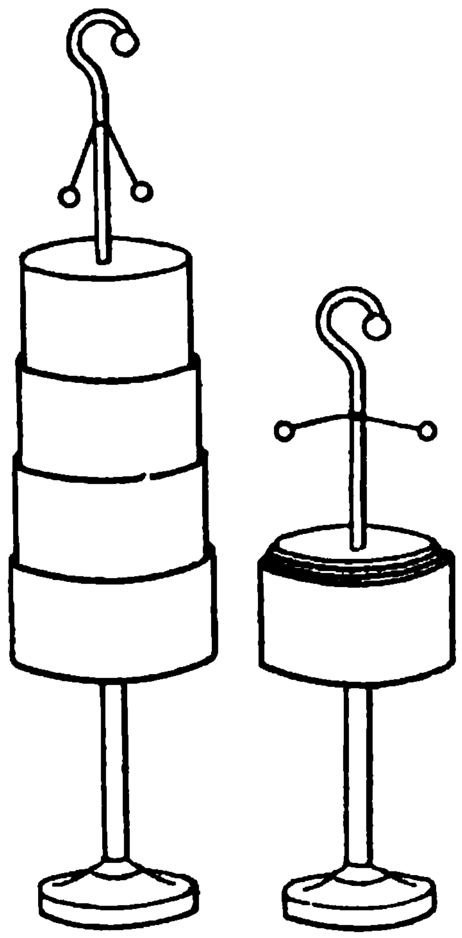


FIG. 280.

The fact that the capacity depends upon the position of neighboring conductors is shown by the apparatus illustrated in Fig. 281. A, an insulated metal disk, is connected by a wire with the electroscope E. The divergence of E indicates as before the potential, V, caused by a charge Q. Now bring up the earthed disk B, and the potential V is lowered as indicated by the convergence of the electroscope. That is, to raise A to the potential V requires a greater charge if B is nearer A; in other words the electrical capacity of A is increased by the presence of the conductor B. If a plate of hard rubber, sulphur, or glass be now put between A and B, the electroscope converges; that is, the

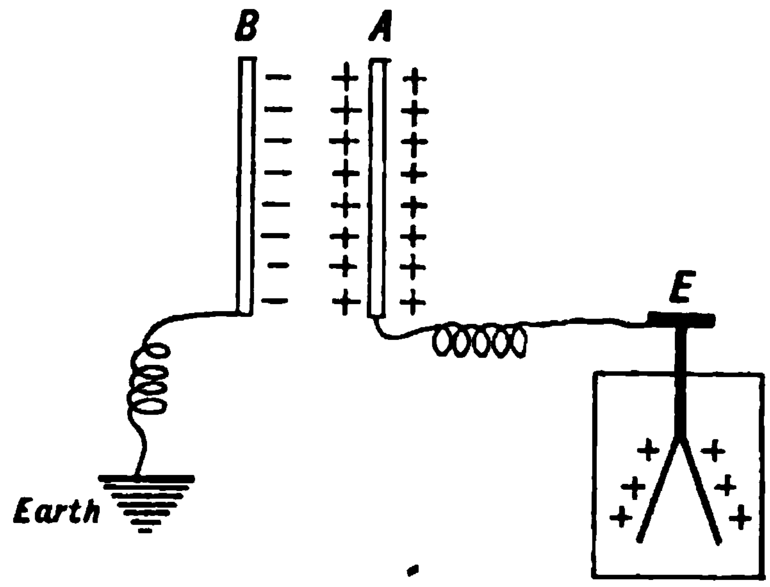


FIG. 281.

potential falls while the charges remain the same. The electrical capacity thus depends upon the intervening medium. We thus see that the electrical capacity of a conductor depends upon (*a*) the size, (*b*) the shape of the conductor, (*c*) the position of neighboring conductors, and (*d*) the intervening dielectrics. The arrangement shown in Fig. 282, consisting of two conductors separated by a dielectric, thus serves to increase the electrical capacity of the insulated conductor, and is called *an electric condenser.*

That the electrical capacity of the conductor A is increased by the presence of the conductor B follows from the definition of electrical potential. The potential of A is equal to the work required to bring a test unit charge from the earth to the body A against the force of the electrical field (Fig. 282). But if we have not only the charge $+Q$, but also the induced opposite charge $-Q$, the force acting on the test unit is less; that is, the work is less, or the potential of A is lowered. Hence the charge to raise A to a given potential must be greater; that is, the electrical capacity of A is increased.

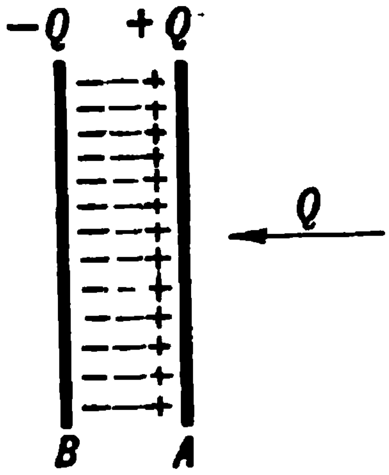


FIG. 282.

411. Condensers.—The arrangement of two conductors separated by an insulator or dielectric is that found in the Leyden jar (Fig. 283). This consists of a glass jar, coated inside and

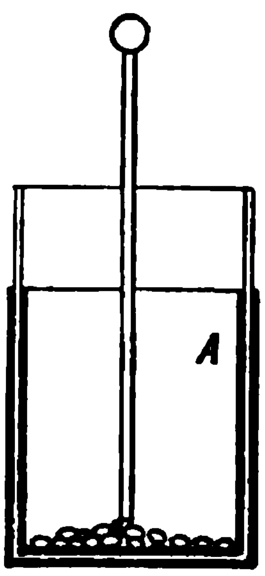


FIG. 283.

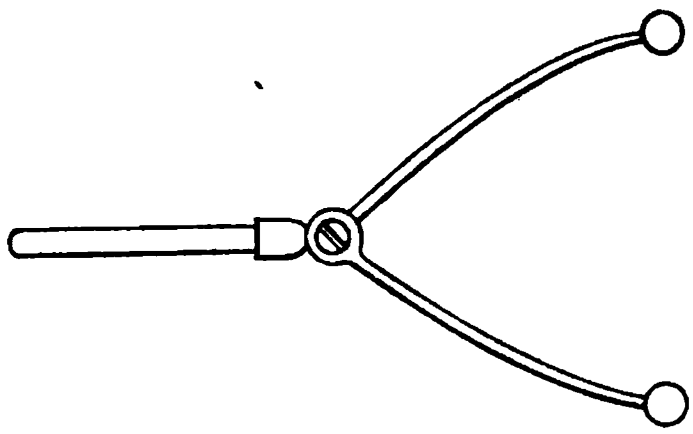


FIG. 284.

outside for about two-thirds of its height by tin-foil. Connection with the inside is made by a supported brass rod terminat-

ing in a ball. When the inside coating A is charged, an opposite "bound" charge is induced on the outside coating B, and the "free" induced charge escapes to the earth. If the inner and outer coatings are connected, there is a spark and an electric discharge which is often very violent. In discharging a Leyden jar, it is safe and convenient to use discharge tongs with an insulating handle (Fig. 284).

Fig. 285.

Very compact condensers are made by piling up sheets of tin-foil separated by sheets of mica or of paraffined paper (Fig. 285). The alternate tin-foil sheets are joined and thus a great capacity is secured in a small space. Mica condensers are used in the laboratory for standards in testing. Fig. 286 shows a common form. The condensers in the bases of induction coils (§511) are generally paraffined-paper condensers.

Fig. 286.

412. Residual Discharge.—A succession of discharges can usually be obtained from a charged Leyden jar or other condenser. Thus, when the inner and outer coatings are connected, there is a brilliant spark, and, if they are connected a half minute later, still another spark discharge takes place; and this may often be repeated a half dozen times, each successive discharge being less than the preceding. These later discharges are called *residual discharges.* The number and magnitude of the residual discharges differ greatly for different condensers, depending upon the kind and thickness of the dielectric. Air condensers show no residual discharges.

The residual discharges are explained as due to the "absorption" of the charges by the dielectric, and the gradual escape of these charges. The "absorption" is, however, probably a state of strain with associated stresses in the dielectric. Just as rubber recovers gradually from distortion (§179), so the dielectric takes

time to recover. Homogeneous dielectrics such as gases and quartz show no residual effects.

The above assumes that the energy of the charged Leyden jar is to be found in the glass, or other dielectric. A very beautiful experiment made by Benjamin Franklin as early as 1748, showed that, in the Leyden jar, "the whole force of the bottle and power of giving a shock is in the glass itself; the non-electrics (conductors) in contact with two surfaces serving only to give and receive to and from the several parts of glass; that is, to give on one side and take away on the other." Franklin used a glass jar coated on the outside with lead foil and having water on the inside for the inner conductor. Connection with the water was made by a wire supported by the cork stopper of the

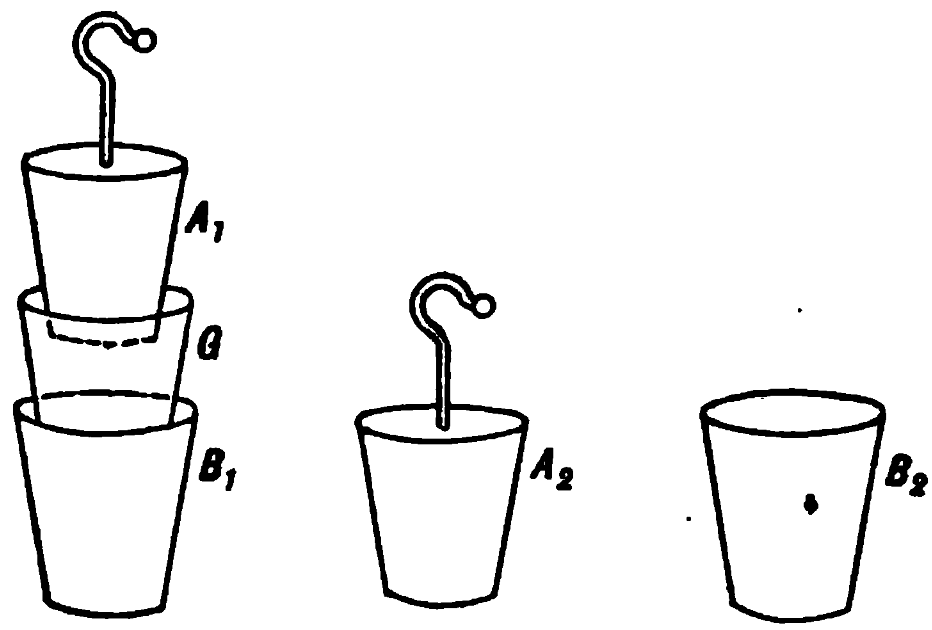


FIG. 287.

bottle. The jar or bottle was charged from an electrical machine. The water was then poured out and found to be uncharged. Fresh water was poured into the bottle, and a bright spark was obtained by discharge of the jar. A common apparatus for Franklin's experiment, known as the separable Leyden jar, is shown in Fig. 287. It consists of a glass cup, G, in which the metal cone A_1 fits as inner conductor, and the metal cup B_1, in which the glass cup G fits. A is charged, and the usual discharge takes place upon connecting A_1 with B_1. If A_1 is again charged, and then lifted out with a rubber handle, and B is removed, it is found that neither A_1 nor B_1 is charged. If now the jar be built up in the reverse order with the same glass cup G, but with a second set of conductors, A_2 and B_2, a bright spark discharge is

obtained from the Leyden jar. From this we conclude that the essential part is the dielectric.

413. Dielectric Properties. Specific Inductive Capacity.—The first person to publish a systematic study of the dielectrics in condensers was Michael Faraday. Faraday used two similar spherical condensers, each consisting of a brass sphere, A, (Fig. 288), suspended by an insulating support at the center of a hollow spherical shell, B. The shell was made of two flanged hemispheres which could be separated so that different dielectrics could be introduced. One of the condensers with air for dielectric was given a charge Q. Its potential, V, was then $V=Q/C$. If this condenser was then joined to the second condenser with air also as dielectric, the charge Q was shared equally between the two, and the potential V was halved. But if the second condenser had another dielectric, say sulphur, the original potential was reduced to V', which was less than half of V. Let C' be the capacity of the second condenser. Then $(C'+C)V'=Q$ and $CV=Q$, and hence $C'/C=(V-V')/V'$. Faraday found that his condenser with sulphur forming part of the dielectric had a capacity 1.6 times the capacity of a similar condenser with air alone as dielectric. Hence, the "inductive action" across sulphur must be greater than across air.

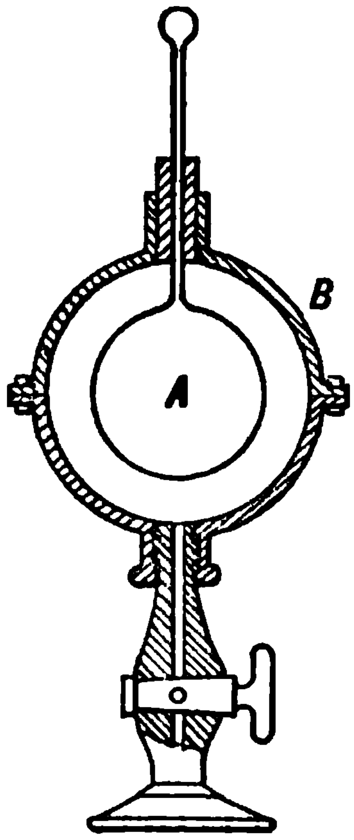


FIG. 288.

Faraday in this way discovered that "different dielectric bodies possess an influence over the degree of induction which takes place through them." He described this by saying that these dielectrics had different "specific inductive capacities." The specific inductive capacity or "*dielectric constant*" of a substance, as it is called more frequently now, is *the ratio of the capacity of a condenser with the given substance as dielectric to the capacity of the same condenser with air as dielectric.* The following is a list of the dielectric constants of some common dielectrics:

Paraffine wax	2.0 to 2.3	Mica	6. to 8.
Petroleum	2.07	Glass	6.6 to 9.9
Hard rubber	2.0 to 3.1	Distilled water	75. +
Sulphur	2.2 to 4.	Alcohol	25. +

These are the same constants as appear in Coulomb's law of electrical force (§401), and one method of obtaining these constants is to measure the force between charged conductors in the different dielectrics. The reason for the connection is readily seen from the remark at the end of §406.

414. Units of Capacity.—A conductor or condenser has unit capacity when unit quantity of electricity raises it to unit potential; that is, C is unity when Q and V are each unity. If Q and V are expressed in c.g.s. electrostatic units of quantity and potential (§405), the above is the definition of the *c.g.s. electrostatic unit of capacity.*

If we use the coulomb and volt as the units of quantity and potential (§441), the corresponding unit of capacity is the *farad.* The *farad* is the electrical capacity of a conductor which requires a coulomb to raise its potential to one volt. From the relations $C=Q/V$; 1 coulomb $=3\times 10^9$ c.g.s. electrostatic units; and 1 volt $=1/3\times 10^{-2}$ c.g.s. electrostatic units; we get 1 farad $=9\times(10)^{11}$ c.g.s. electrostatic units of capacity. The *microfarad,* the millionth of a farad, is the ordinary practical unit of capacity. From the above we see that a microfarad $=9\times(10)^5$ c.g.s. e.s. units.

415. Capacity Calculations.—The capacity of certain forms of conductors and condensers can be calculated when the dimensions of the parts, and the dielectric constant of the insulator are known. We give here the formulæ for the capacity C of a sphere and of three forms of condensers. The dielectric constant K is unity if air is the insulating medium.

For an isolated sphere, with radius r, in air,

$$C=r$$

For at all points inside the sphere the potential is the same and equal to that at the surface of the sphere. Since all parts of the charge on the surface of the sphere are at the same distance from the center, the potential at the center is $V=Q/r$ or $Q/V=r$. Hence the capacity of the sphere is equal to r.

For a sphere of radius r surrounded by a concentric spherical shell of internal radius r' (Fig. 289).

$$C=K\frac{rr'}{r'-r}$$

For, if the charge on the inner sphere is Q, that on the outer sphere is $-Q$. The outer sphere, being connected to the earth, is at zero potential. The potential at any point on the surface of the inner sphere due to the charge on the outer sphere is $-Q/r'$ when the dielectric is air and that due to the charge on the inner sphere is Q/r. Hence the potential of the inner sphere is $V = Q(1/r - 1/r')$ etc.

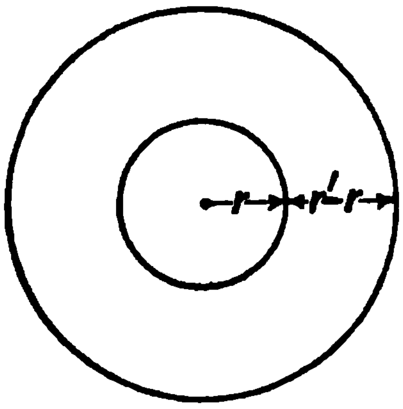


FIG. 289.

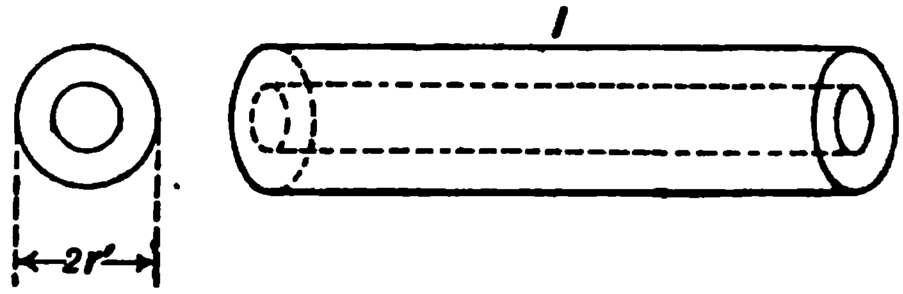


FIG. 290.

For a cylinder of radius r surrounded by a concentric cylindrical shell of internal radius r' (Fig. 290), and of length l, large compared to r and r'

$$C = K\frac{l}{2\log_e\frac{r'}{r}}$$

For a pair of equal parallel plates of area A, a relatively small distance, d, apart

$$C = K\frac{A}{4\pi d}$$

The proofs of these formulæ can be found in the more extended treatises on electricity.

416. Energy of a Charged Conductor.—The discharge spark with its accompanying light and heat is sufficient evidence that a charged conductor or condenser has energy. An expression for this energy is most easily found by calculating the work required to charge the body. Let Q equal the charge required to raise the body to the potential V. The potential of a body has been defined as the work required to bring a positive test unit from infinity to the body. The definition assumes that the potential of the conductor is not appreciably changed by the

addition of the test unit. Here, however, we, have to determine the work necessary to bring a charge Q up to a body which is initially at zero potential, but which is raised to the final potential V by the charge Q. We can think of the charge as being brought in a very large number of small fractional charges Q/n. The potential of the body will be raised equal amounts as each fractional charge is added, until the final potential V is reached. The average potential during the charge is thus

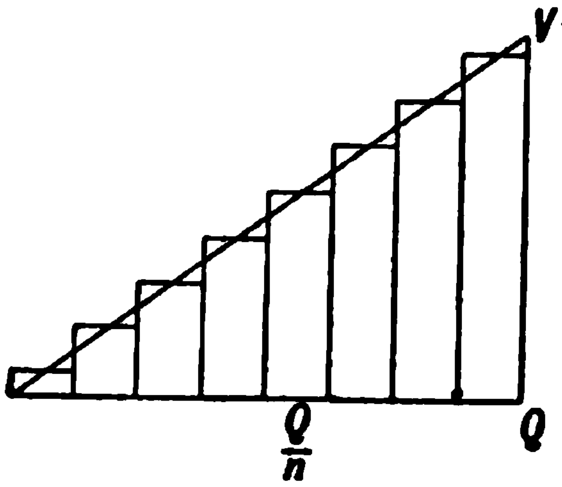


FIG. 291.

$V/2$, and hence the total work equals the product of the total charge Q and the average potential, and so the energy $E=QV/2$. The process can also be represented graphically (Fig. 291) and the result obtained from the area of a triangle as in the similar cases considered in §§27, 56. Since $Q=CV$ (§410), we can write this in the form, $E=CV^2/2$.

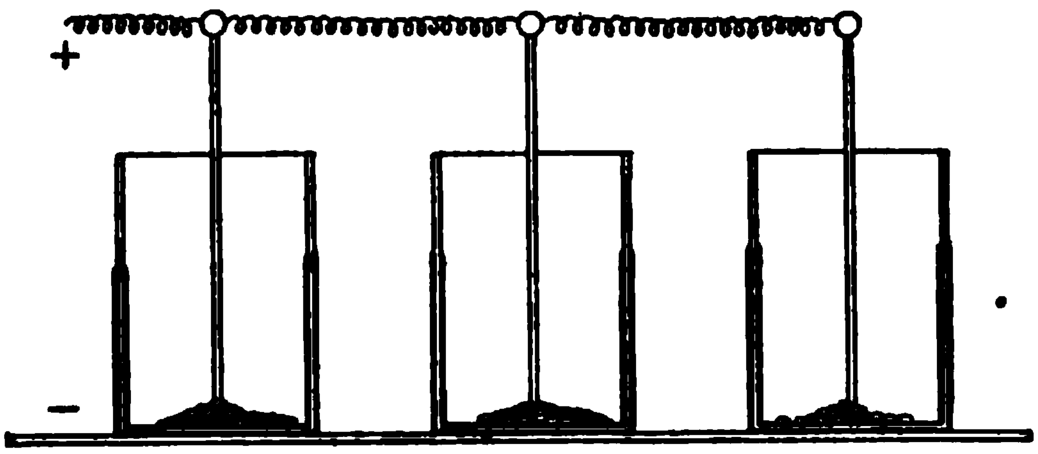


FIG. 292.

417. Condensers "In Parallel" and "In Series."—Condensers can be joined together "in parallel" or "in series." When joined in parallel, all the positive coatings are connected to form one terminal and side of the battery of condensers, and all the negative coatings are connected to form the other terminal and side (Fig. 292). The resultant capacity is that of a single large condenser, with coatings equal to the sum of the coatings of the individual condensers, or $C=C_1+C_2+C_3+$, etc

Condensers are joined, "in series" (or "in cascade") when they are insulated, and the outer coating of the first is connected to the inner coating of the second and the outer coating of the second is

joined to the inner coating of the third, and so on through the series of condensers (Fig. 293). To calculate the capacity of the battery "in series," let V_1, V_2, V_3, etc., be the potentials and C_1, C_2, C_3, etc., the capacities of the condensers. The

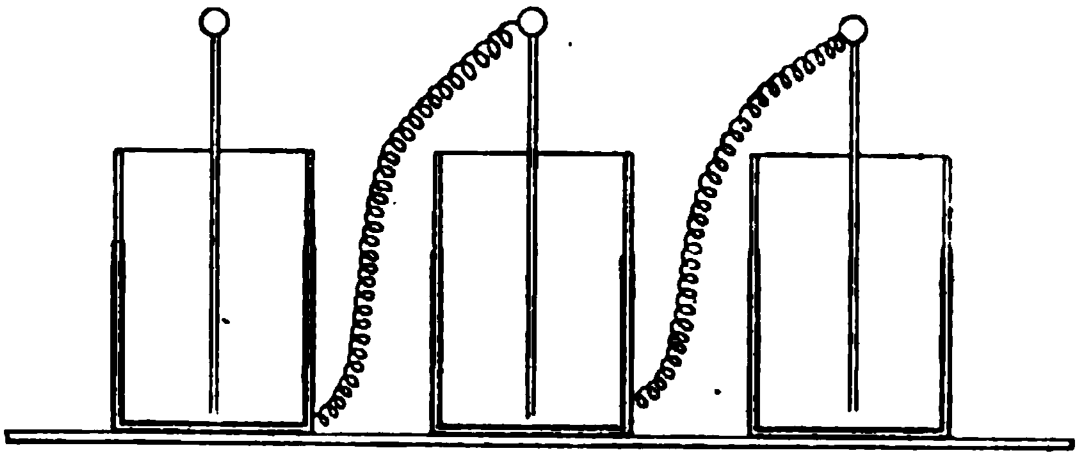


Fig. 293.

quantity Q on each inner coating will in each case equal the quantity on the outer coating, and this will equal the quantity on the next inner coating since they are corresponding induced charges. Thus

$$Q=(V_1-V_2)C_1=(V_2-V_3)C_2=(V_3-V_4)C_3=C(V_1-V_4)$$

where C is the resultant capacity of all the series. Hence

$$\frac{Q}{C}=\frac{Q}{C_1}+\frac{Q}{C_2}+\frac{Q}{C_3}$$

and

$$C=\frac{1}{1/C_1+1/C_2+1/C_3}$$

418. Quantitative Use of Lines and Tubes of Force.—In representing the state of an electric field by lines of force, only the question of the direction of the force has been considered. To represent the magnitude of the force, we limit the number of lines of force from a body, so that the number of lines is equal to the number of units of positive charge. That is, if there are q units of electricity, there will be q lines of force. Each line is then called " a unit line," or simply a "line of force."

Suppose the charge q to be at a point, and distant from other charges. The lines are then radial and symmetrical about q. Draw a sphere with radius r and with q at the center. Through this spherical surface, there are $N=q/4\pi r^2$ lines per square centi-

meter. But the force on unit charge at adistance r is $F=q/Kr^2$ (§401) hence $F=4\pi N/K$, where N is the number of unit lines through a square centimeter section taken normal to the field. But the force F acting on unit charge at a point is called *the intensity of the field at the point.* The intensity of the field at a point is thus equal to $4\pi/K$ times the number of unit lines per square centimeter of normal section of the field.

For the more complete representation of the state of an electric field, we introduce the conception of "tubes of force." A tube of force is a channel bounded by lines of force and having, as one end, the area covered by a positive charge, and, as the other end, the area covered by the corresponding negative charge. A "unit tube" or a "Faraday tube" as Professor J. J. Thomson calls it, has unit positive charge at one end, and unit negative charge at the other end. Hence as many unit tubes start from a body as the body has units of positive electricity on it. The electric field is thus filled with these tubes of force.

The terms "unit line" and "unit tube" are thus equivalent. We see that the intensity at a point of a field is equal to $4\pi/K$ times the number of tubes of force per square centimeter of section normal to the field, or $F=4\pi N/K$. If there are N tubes across a square centimeter, then s, the cross-section of each tube, is $s=1/N=4\pi/FK$. Hence $Fs=4\pi/K$, or the product of the cross-section of a tube and the intensity of field is the same at every section of the tube. We thus see that where the intensity decreases the tube widens out, and where the intensity increases the tube narrows. This suggests the flow of a stream, and hence Maxwell and others use the term "flux" or "flow" of force, for the quantity Fs.

419. Energy of an Electric Field.—To calculate the energy in unit volume of an electric field, take the case of two parallel charged plates, with a difference of potential V and a charge of Q. Q unit tubes extend between the two plates. The total energy is $(VQ/2)$ (§416), and so the energy per tube is $V/2$ ergs. If the tube has a length L, the energy per unit length of tube is $V/2L$. But $V=FL$ by the definition of difference of potential and so the energy per unit length of the tube is $F/2$. We have already seen that the number of tubes across a normal square centimeter is $N=FK/4\pi$ (§418). But N tubes each of unit length, occupy a cubic centimeter. Hence the energy per cubic centimeter of the medium is $F^2K/8\pi$.

420. Atmospheric Electricity.—In 1752 Benjamin Franklin described the famous kite experiment by which he "completely demonstrated the sameness of the electrical matter with that of lightning." A silk kite on which there was a pointed wire was raised, and it was found that "as soon as any of the thunder-clouds came over the kite, the pointed wire drew the electric fire from them, and the kite with all the twine became electrified." By this means Franklin got electric sparks and also was able to charge a Leyden "phial" from the clouds.

The next marked advance in the study of atmospheric electricity was due to the invention and use of the water-drop electrograph and the electrometer by Kelvin about the middle of the last century. Kelvin showed that the end of a tube from which a stream of water breaks into drops takes the electric potential of the air at the point. It was found that the potential of the air in dry weather is normally positive relative to the earth and increased with the height. The potential gradient is expressed in volts per meter rise in height. This may be several hundred volts per meter, but it varies greatly with the season, the time of day and the weather conditions. It is also not always positive, for the potential of the atmosphere at times is negative relative to the earth, and is very frequently so in rainy weather. The causes of atmospheric electricity are not definitely determined. Evaporation, friction of the clouds, the action of ultra-violet light, and of radio-active materials are some of the causes suggested.

The electrical phenomena of the atmosphere more commonly observed are forms of lightning and the aurora borealis. Lightning is an electric discharge between clouds, or between the clouds and the earth. It takes the form of forked lightning, sheet or "heat" lightning and "ball" lightning. The "ball" lightning is not well understood and may be due to an optical illusion. Recent experiments by the resonance methods (§542) of electrical waves show that lightning discharges are oscillatory (§541).

One of the first applications of electrical science was Franklin's use of points, "lightning rods," for the protection of buildings against injury by lightning discharges. The protection from lightning rods is probably greater in the way of silently dis-

charging the surrounding atmosphere, rather than in conducting away disruptive discharges. The best protection against lightning is a metallic net work covering the building more or less completely and having a good connection to moist earth.

The aurora borealis or "northern lights," is an electrical discharge in the upper atmosphere, and is most frequently seen towards the polar regions. It is thought to be analogous to the electrical discharges in vacuum tubes.

ELECTROKINETICS

421. The Electric Current.—If two conductors *A* and *B*, charged to different electrical potentials, be connected by a long thin wire, there will be a flow of electricity, that is, an electric current, through the wire. The direction of this current is defined as being from the higher to the lower potential. The current will continue so long as there is a difference of potential between the ends of the wire. In the case of electrostatic charges, such as those of the Leyden jar, the potentials are equalized in a very small fraction of a second, that is, the currents are momentary. (We shall see later in §§541 and 542 that under frequent conditions, especially if the connecting wire is short and thick, they are also oscillatory.) To get a continued current in a wire, the difference of potential must be kept up. The potentials produced by electrostatic machines are very high and the electric quantities separated are small, so that the currents from such machines are small, momentary and intermittent. For producing and maintaining continued electric currents, voltaic cells, thermo-couples and dynamo-electric machines are used.

These will be described later, but it will be convenient to state at this stage the principle of the voltaic cell, since it was the first device discovered for obtaining continued currents and voltaic cells of some form are usually employed for most of the experiments which we shall presently describe.

422. The Voltaic Cell.—When two dissimilar conductors, *A* and *C* (Fig. 294*a*) are immersed in a liquid, *B*, which acts chemically on at least one of them, and the parts out of the liquid are connected by a wire, *D*, a current of electricity flows in the wire, heating it and producing other effects to be described

presently. If the wire D be cut and its free ends be joined to a sufficiently sensitive electrometer (§407) the latter will show by the deflection of its needle that the ends of the wire are at different potentials. Volta, to whom we owe the above discoveries, accounted for this difference of potential by assuming that there is an abrupt difference of potential set up at the contact of each pair of dissimilar conductors in the circuit. This view has been generally accepted but much difference of opinion has existed as

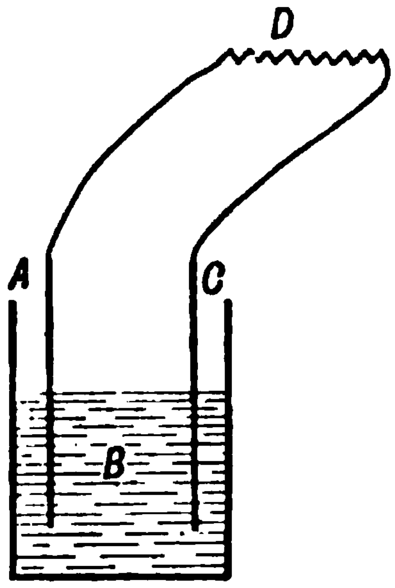


FIG. 294a.

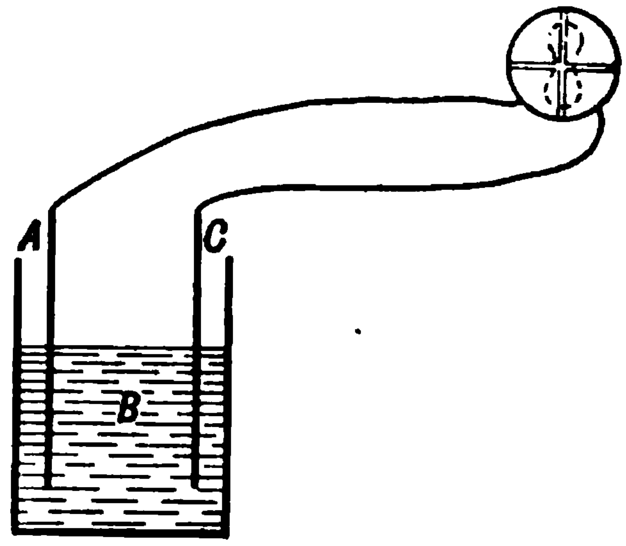


FIG. 294b.

to the relative magnitudes of these differences of potential at the various contacts, D with A, A with B, B with C, and C with D. It will not be necessary to consider this controversy further at present (see §475).

423. Electromotive Force.—Assuming that there are such contact differences of potential, let us denote the *rise* of potential from D to A (positive or negative) by V_{DA} and so on for the other contacts in the circuit. Then the whole difference of potential of the ends of D connected to the electrometer and measured by it is $V_{DA}+V_{AB}+V_{BC}+V_{CD}$. This, by the definition of potential difference (§402), is the work that would be done in taking a unit quantity of electricity from one free end of D to the other along the line of the conductors in the circuit. When the ends of D are joined and a current flows, we may regard the current as being due to the sum of the steps of potential at the contacts and this sum is accordingly called the electromotive force, E, acting in the circuit. It is evidently also equal to *the work that would be done in taking a unit quantity once around the circuit.* The latter is the general measure of the electromotive force in a

circuit, whether it be due to voltaic cells, thermoelectric junctions, or a dynamo, in the circuit.

424. Two Classes of Conductors.—Volta also sought to obtain currents by circuits consisting of metallic conductors only, but in this he did not succeed. He found, in fact, that, whatever differences may be supposed to exist at the various junctions in such a circuit, the sum of these formed as described in §423 is equal to zero, that is, the electromotive force produced in such a circuit is zero. (We now know that such is not the case if there are differences of temperature in the circuit.) He was, therefore, led to divide conductors into two classes, conductors of the *first class* being such as are not competent by themselves to produce an electromotive force when joined in a circuit, at least one conductor of the *second class* being necessary for a finite electromotive force. The former class includes all metallic conductors, while the latter, now called electrolytes (§462), are chemical compounds which can be decomposed by an electric current.

It follows from the above that if we suppose the liquid of the voltaic cell (Fig. 294) to be absent and A and C to be directly in contact, then $V_{DA}+V_{AC}+V_{CD}=0$. We may also evidently write this equation in the form $V_{CD}+V_{DA}=V_{CA}$, or the contact rise of potential from C to A, if placed directly in contact, is equal to the sum of the rises of potential from C to D and from D to A, a result that holds for any three conductors of the first class.

425. Electromotive Force of a Cell.—While an electromotive force consists in general of parts that are located at different points in a circuit and is measured by the work done in taking unit of electricity once around the circuit, it may, in the case of a voltaic cell, be considered as due solely to the liquid and the plates directly in contact with it. For, from the formulas stated in §§423, 424, it follows that

$$\begin{aligned} E &= V_{AB}+V_{BC}+V_{CD}+V_{DA} \\ &= V_{AB}+V_{BC}+V_{CA} \end{aligned}$$

This could not, however, be used as the basis for a satisfactory practical definition of the electromotive force of a cell, since, as has been stated, there is some doubt as to the parts that the separate terms contribute to the whole sum. Of this whole sum

there is no doubt, since it can be measured directly by an electrometer as stated in §422. We shall, therefore, define the e.m.f. of a cell as follows:

The electromotive force of a voltaic cell is the difference of potential of wires of the same material connected to the plates of the cell, when the cell is an open circuit.

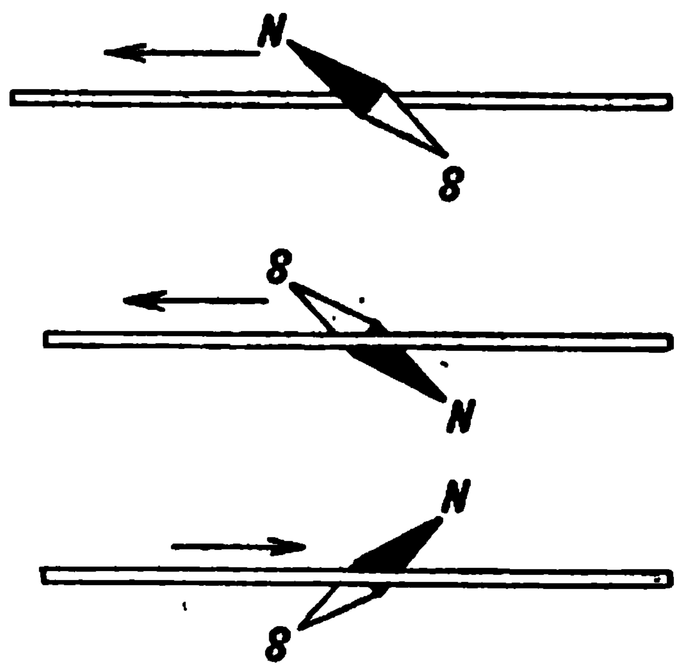


Fig. 295.

When the circuit is closed by metallic wires of any kind, the electromotive force of the circuit is, as we have seen above, equal to that of the cell and independent of the material of the connections (provided they be all at the same temperature).

426. Magnetic Effect of an Electric Current.—In 1820 Hans Christian Oersted, Professor in the University of Copenhagen, made the epoch-making discovery that an electric current acted on a neighboring magnetic needle. It was found that when a straight wire was held in a north and south line over a magnetic needle, the needle was deflected if an electric current was passed through the wire. Further, the direction of the deflection of the needle was reversed, (*a*) by reversing the direction of the current, and (*b*) by holding the wire under, instead of over the needle. This is illustrated in Fig. 295.

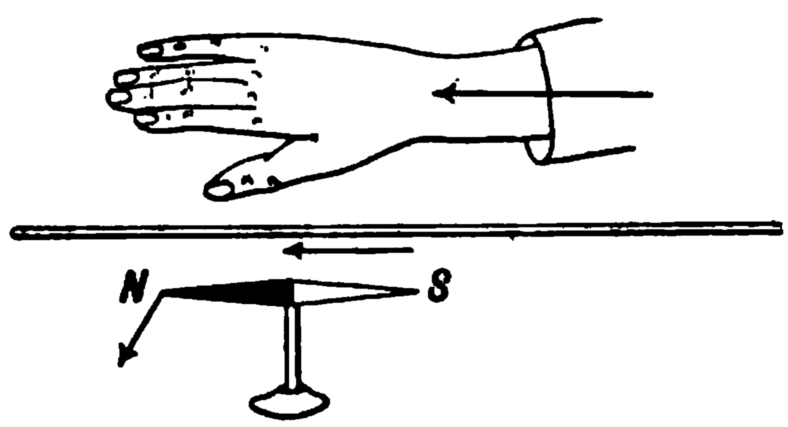


Fig. 296.

The directions of the currents and of the corresponding deflections are described by the following rule: Hold the open right hand on the side of the wire opposite the needle, with the palm toward the needle, and the fingers pointed in the direction of the current, then the thumb indicates the direction of the deflection of the *N* pole of the needle (Fig. 296).

Oersted's experiment shows that a wire carrying an electric current is surrounded by a magnetic field, and that the direction

of the field is on all sides perpendicular to the current direction, that is, that the magnetic lines must be circles about the current. This can also be shown by means of iron filings. A vertical wire passes through a hole bored in a horizontal glass plate; if a strong current is passed through the wire, and iron filings are sprinkled on the glass, it is seen that the filings arrange themselves in circles with the wire as center (Fig. 297). By using a small compass, it is easy to fix the direction of this field. The direction of the current and that of the accompanying magnetic field is stated by Maxwell's rule: *If the direction of the current is that of the advance or thrust of a right-handed screw, then the direction of rotation of the screw gives the direction of the magnetic field.* This is illustrated in Fig. 298.

FIG. 297.

From the above we can see, that a N pole would rotate in a circle about a current, provided the N pole could be isolated from its S pole. Fig. 299 shows a piece of apparatus for demonstrating this rotation and its direction. The current from the battery

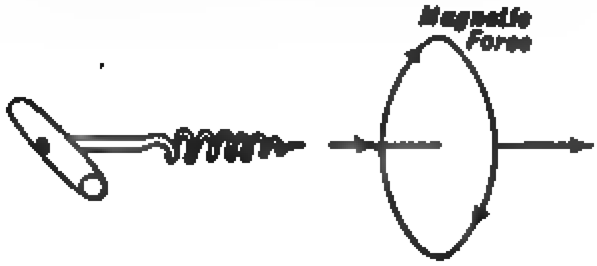


FIG. 298.

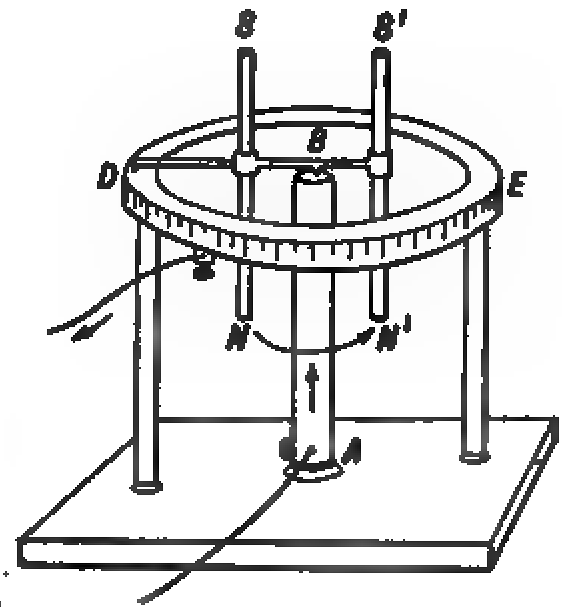


FIG. 299.

enters from below at A, passes up the vertical rod to B, and by a mercury cup enters the horizontal arm BD; by this it reaches the circular mercury trough E, and completes the circuit back to the battery. Suspended so

that they can rotate with the arm *BD*, about the axis of the vertical current *AB*, are the two magnets *NS* and *N'S'*, which have their north poles *N* and *N'* in the field of *AB*, while the two south poles *S* and *S'* are outside of this field. When the current flows from *A* to *B*, the north poles revolve anti-clockwise about the current as looked at from above, and continue to revolve so long as the current continues. Reversing the direction of the current reverses the direction of the rotation.

427. Magnetic Lines of a Circular Circuit and of a Solenoid.—When the wire carrying a current is bent into a circle, as shown in Figs. 300*a* and *b*, the lines of magnetic force pass through the

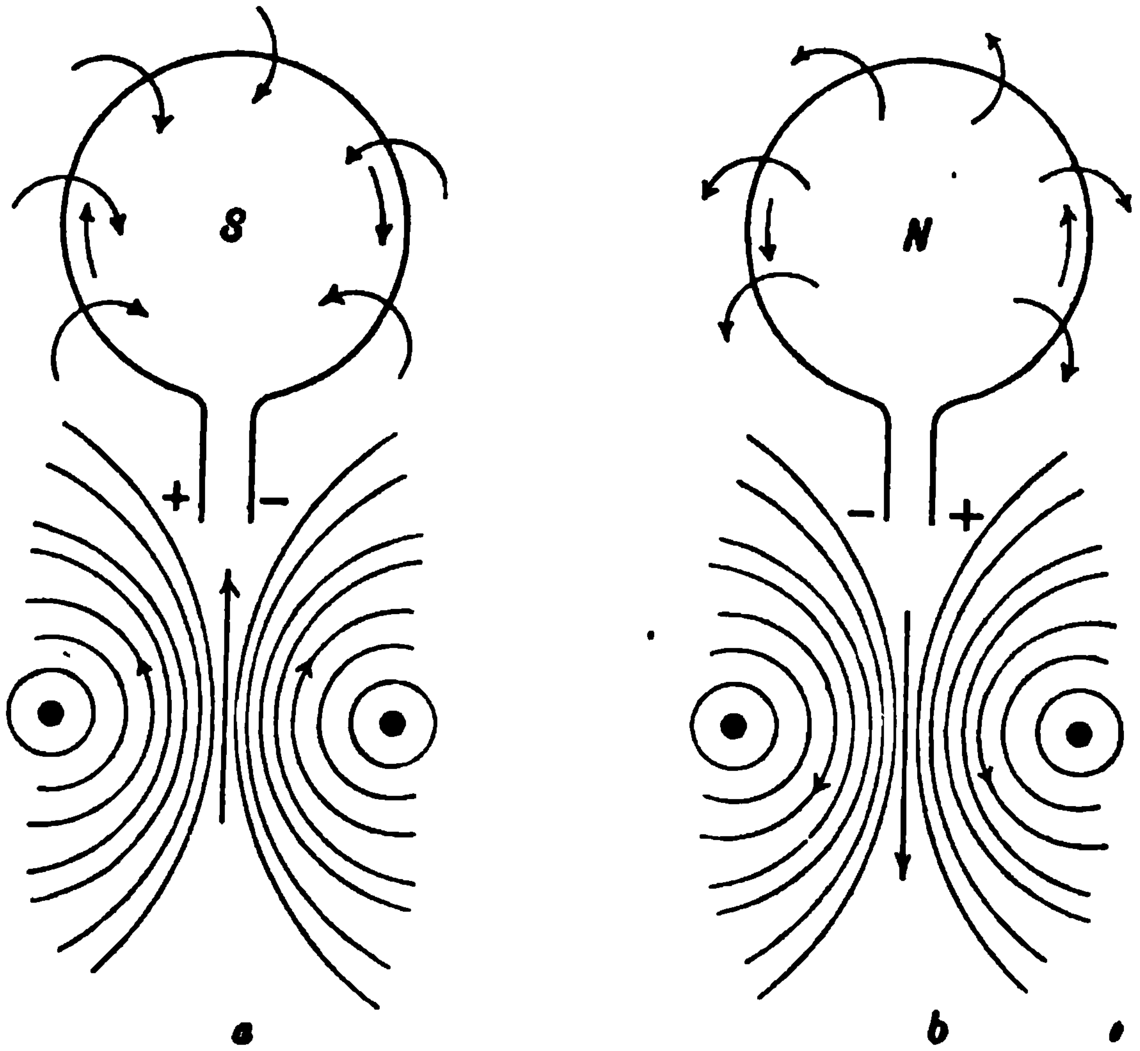


FIG. 300.

area bounded by the circle, entering at one face of the circle and going out from the opposite face. That is, the north pole of a magnetic needle would act as if repelled by one face of the circle and attracted by the other face. The circular circuit then acts like a thin sheet magnet, or "a magnetic shell," one face of which is a north "pole" and the other a south "pole." It is also seen that the direction of the magnetic lines (*S* to *N*) of the shell is related to the direction of the current in the coil as the thrust to the twist of a right-handed screw.

By winding the wire closely on a cylinder in one or more layers, we get a helix or solenoidal coil. It can be considered as a series

of parallel and equal circles with centers on the axis of the cylinder. A helix with a current through it forms a *solenoid*. The magnetic field of a solenoid is indicated in Fig. 301. It is the resultant of the magnetic fields of the individual circular currents. It is seen that one end forms a N pole, and the other end a S pole. The magnetic field inside the solenoid is uniform except near the ends (§430).

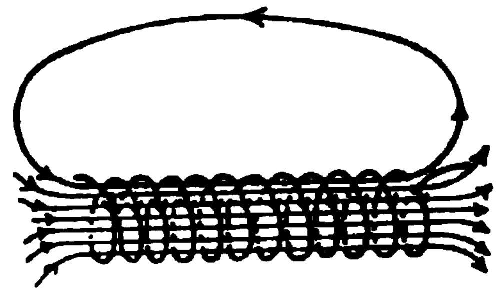


FIG. 301.

428. Current and Field Strength. Units of Current.—The strength of the magnetic field at a point P, due to a current i in a small element of circuit ds at a distance r from P, (Fig. 302a) varies directly as the current i, directly as the length ds resolved at right angles to r, and inversely as the square of the distance r; that is, H, the strength of the magnetic field at P, is given by the equation

$$H = k\frac{ids}{r^2},$$

where k is a constant the value of which depends upon the surrounding medium and upon the units used. The magnetic field at P is evidently at right angles to the plane of Pds. If

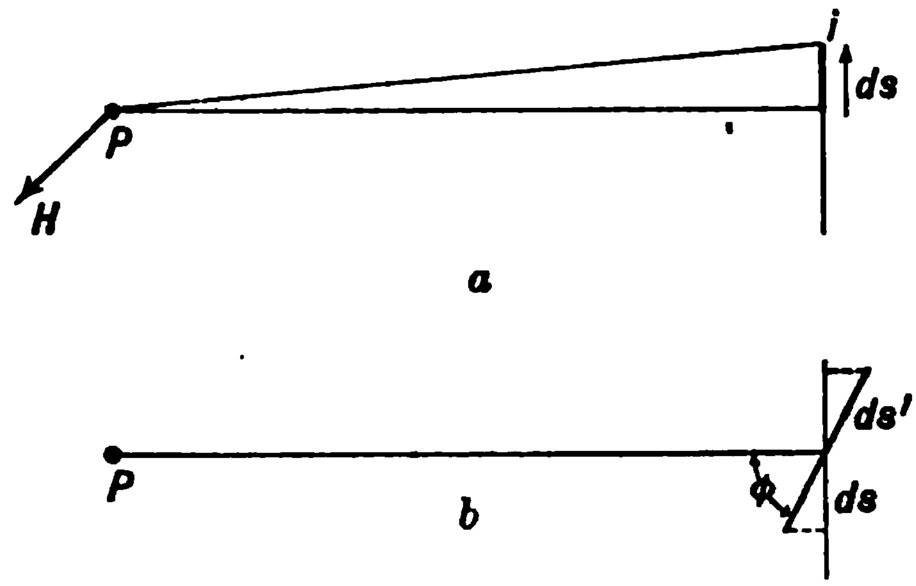


FIG. 302.

ds' is the length of the current element, and if this makes the angle ϕ with r, (Fig. 302, b) then $ds = ds' \sin \phi$ and

$$H = \frac{kids'}{r^2}\sin \phi$$

It is evident that a direct experimental proof of the above law is not possible, since we cannot have an isolated circuit element ds.

We can, however, apply the law to various circuits and deduce formulæ which can be tested.

The simplest application is in calculating the magnetic field at the center of a circular circuit. Here the sum of all the elements ds is $2\pi r$, the circumference of the circle, and all the elements are at the distance r from the center, and hence the magnetic field at the center is

$$H = k\frac{2\pi ri}{r^2} = k\frac{2\pi i}{r}$$

In this equation, the constant k is unity, if we measure r the radius in centimeters, and define unit current as follows: *Unit current is the current which, flowing through a circle of one centimeter radius in air, exerts a force of 2π dynes on a unit magnetic pole at the center of the circle.* This is *the c.g.s. electromagnetic unit of current.* It can also be stated as follows: The electromagnetic unit of current is that current which, flowing through unit length of arc with unit radius, produces unit magnetic field at the center.

If the circular circuit has n turns instead of one turn. the formula becomes,

$$H = \frac{2\pi ni}{r}$$

By sending the same current through circular circuits of different radii and measuring the magnetic fields at the centers, it is found that the above formula holds. It follows from this that the law of action of each short element of a current must also be true. For practical measurements the unit of current used is the *ampere.* The ampere is one-tenth (10^{-1}) of the c.g.s. electromagnetic unit of current.

429. Electromagnetic Unit Quantity of Electricity.—The above definitions of unit current are founded entirely on the magnetic action of a current. In stating them we have implied nothing as to the nature of electricity itself, the direction in which it flows, or the amount that flows. If, however, we now *assume* that an electric current may be regarded as the flow of an incompressible fluid in a definite channel, we must suppose that, as in the case of water flowing in a pipe, the quantity that flows through every cross-section is the same and we are thus led to an entirely new definition of a unit quantity of electricity. *Unit*

quantity of electricity is that quantity which, in each second, passes through every cross-section of a linear conductor which carries a unit steady current.

It should be noted that nothing yet stated enables us to decide whether a current of electricity consists of a flow of positive electricity from high to low potential in a conductor, or a flow of negative electricity in the opposite direction, or a combination of both. This question can only be decided by considerations that will be referred to later. Hence, the total quantity which passes a section of the circuit in t seconds, when current i flows, is $q=it$.

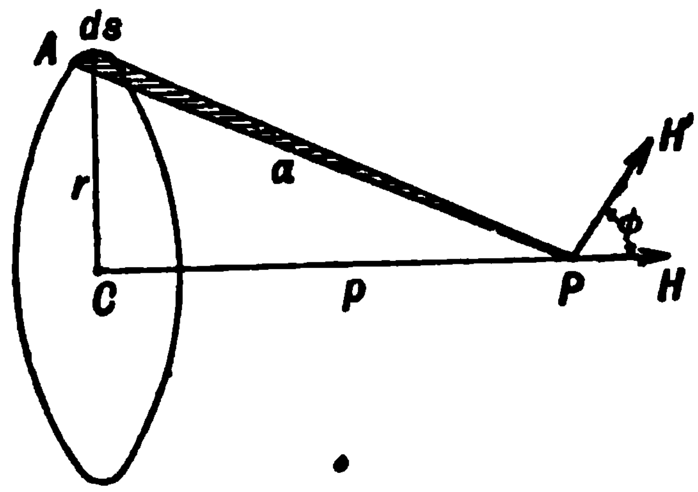


FIG. 303.

The c.g.s. electromagnetic unit of electricity is the quantity carried in a second past a point in a circuit by the c.g.s. electromagnetic unit of current. Experiments show that this is about 3×10^{10} times greater than the c.g.s. electrostatic unit of electricity (§401). One-tenth (10^{-1}) the c.g.s. electromagnetic unit quantity is called a *coulomb*. The coulomb is thus the quantity of electricity which passes any section of a circuit in a second when an ampere flows—hence coulombs = amperes × seconds.

430. Magnetic Field on the Axis of a Circular Circuit. Same for a Solenoid.—The magnetic field at a point P on the axis of a circular circuit can be found as follows: Let p be the distance of P (Fig. 303) from C the center of the circle, and r the radius of the circle. The distance of P from an element ds of the circuit is then $a=\sqrt{r^2+p^2}$, hence the magnetic force on unit pole at P due to current i in ds is

$$H'=\frac{ids}{r^2+p^2}$$

This force is at right angles to the plane P–ds. It is resolved along the axis by multiplying by $\cos\phi=\sin CPA=r/a=r/\sqrt{r^2+p^2}$; or

$$H=\frac{rids}{(r^2+p^2)^{3/2}}$$

For the whole circle the intensity of the magnetic field is therefore

$$H=\frac{2\pi r^2 i}{(r^2+p^2)^{3/2}}$$

It is evident that the component of the force at right angles to the axis for

each element ds will be annulled by that of the element at the opposite end of the diameter, and so the above gives the total field at P.

To get the intensity of the field at a point P on the axis of a solenoid, the action of all the parallel circular circuits must be added. Consider a small section $MM' = dx$ of the solenoid, x being the distance of its center, C, from P. Denote the number of turns in the solenoid by n, its length by L and the strength of the current by i. The number of turns in the length dx is

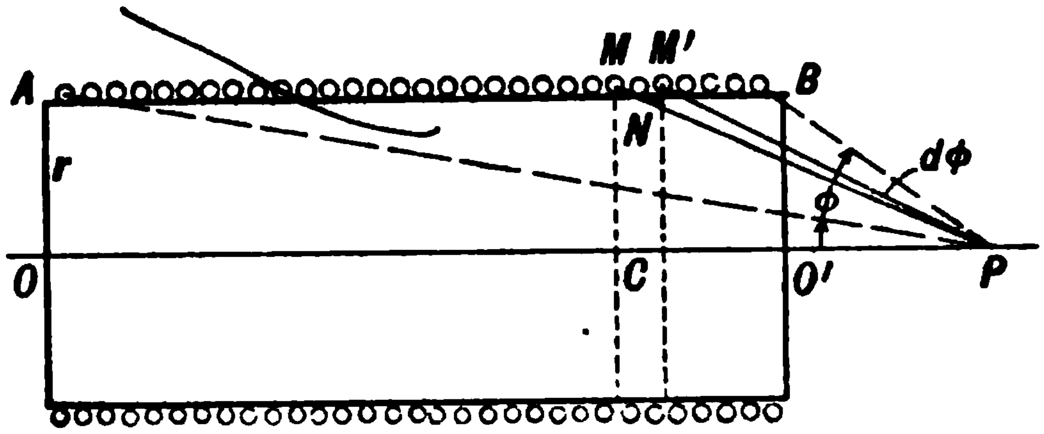


FIG. 304.

ndx/L. We thus get for the intensity of the magnetic field at P due to this element MM' of the solenoid

$$dH_P = \frac{2\pi r^2 nidx}{L(r^2+x^2)^{3/2}}$$

Substituting $x = r \cot \phi$ and $dx = -rd\phi/\sin^2 \phi$ the expression becomes

$$dH_P = -\frac{2\pi ni \sin \phi d\phi}{L}$$

Integrating between the limits $\phi_1 = BPO$ and $\phi_2 = APO$ we find for the total intensity at P

$$H_P = \frac{2\pi ni}{L}(\cos \phi_2 - \cos \phi_1)$$

Now if the length L of the solenoid is large compared to the radius r, and we take the point P near the middle of the solenoid, we can put $\phi_1 = 180°$ and $\phi_2 = 0°$, and hence

$$H_P = \frac{4\pi ni}{L}$$

This formula also gives (approximately) the field intensity inside a ring solenoid.

It is to be noted that at the ends of a straight solenoid, where ϕ_1 or $\phi_2 = 90°$, we have $H_P = 2\pi ni/L$.

431. Field About a Straight Circuit.—The intensity of the magnetic field due to a current in a straight circuit of indefinite length varies inversely as the distance of the point from the circuit. This can be proved experimentally by the arrangement shown in Fig. 305. AOB is a vertical circuit, and NS is a magnet placed upon a horizontal disk which is free to rotate about the

circuit as axis. If r_1 and r_2 are the distances of the poles $+m$ and $-m$ from the center O, and H_1 and H_2 the intensities of the field at the two poles, the moments of the force about O are mH_1r_1 and $-mH_2r_2$. Experiment shows that there is no rotation, that is, that the moments of force are equal and opposite or that $mH_1r_1 = mH_2r_2$, and hence $H_1/H_2 = r_2/r_1$. The intensities of the field due to the circuit are, therefore, inversely as the distances r_1 and r_2. This is known as Biot and Savart's law.

From this we get that $H = ki/r$, for the field about a straight

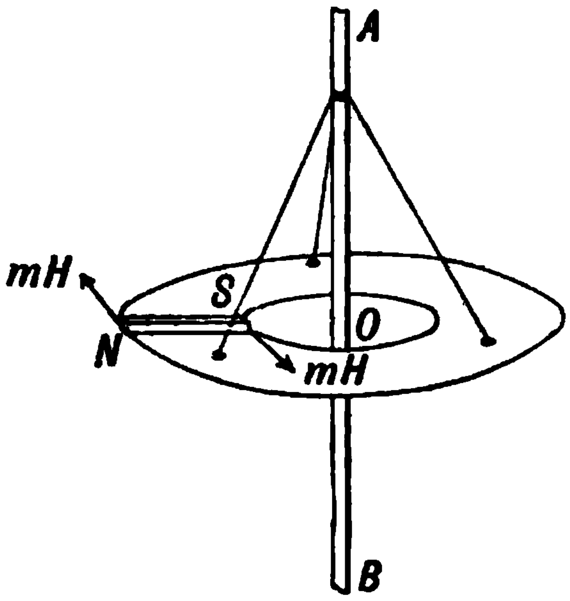


FIG. 305.

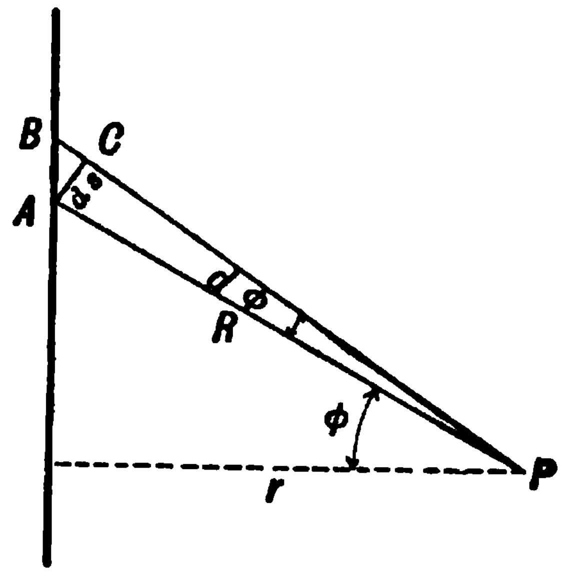


FIG. 306

circuit of indefinite length. It can be shown by mathematics and also by experiment, that $k=2$, if i is measured in c.g.s. e.m. units of current, that is, $H = 2i/r$.

The mathematical proof is as follows: The effect at P (Fig. 306) of the element AB of the current is the same as that of its projection, ds or AC, perpendicular to R. Now $AC = Rd\phi$ and $r = R \cos \phi$. Hence (see §428) $ids/R^2 = i \cos \phi d\phi/r$. Integrating this between limits $-\pi/2$ and $+\pi/2$ we get $H = 2i/r$.

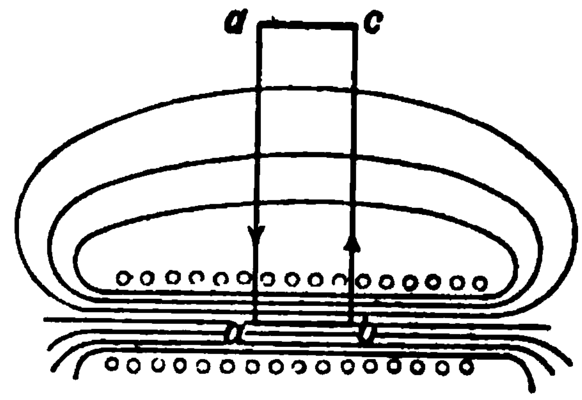


FIG. 307.

We can now calculate the work done in carrying a pole m round a current i. For $W = Hml = \frac{2im}{r} \cdot 2\pi r = 4\pi im$; or, when a unit pole moves about a circuit which carries a current i, $4\pi i$ ergs work are done.

From the above we get an instructive proof of the intensity of the field at the center of a solenoid. Take a closed path $abcd$ (Fig. 307). The side ab is a straight line parallel to the lines of force. The sides bc and ad are perpendicular to the field and extend

indefinitely, that is, to a region where the field becomes null. Then in moving a pole m around the path $abcd$, the only force is along the line ab. Hence the work $W = H.m.ab$, where H is the intensity of the field along ab. If n is the number of turns in the solenoid and L its length, the number of turns in the length ab is $n.ab/L$, and if i is the current, $W = 4\pi mi.\ ab.\ n/L$. Hence $H = \frac{4\pi ni}{L}$. This formula holds true approximately for any point within a long solenoid and not near either end (§430).

432. The Electric Current and the Magnetic Field.—Oersted's discovery shows that an electric current is not simply a transfer of electricity along or in a conductor, but that the whole region about the conductor is involved. With the transfer of electricity there is a magnetic field at right angles to the same. Professor J. J. Thomson has shown how the various phenomena of the electromagnetic field may all be interpreted as due to the motion of the electric lines or "Faraday tubes" (§418), which accompanies the transfer of the electric charges. Suppose we have two

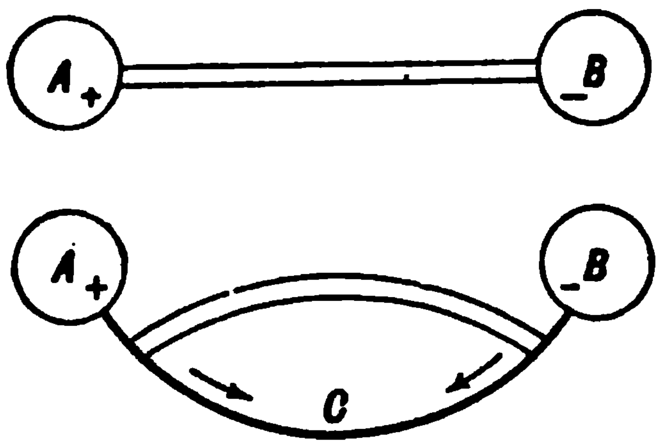


FIG. 308a.

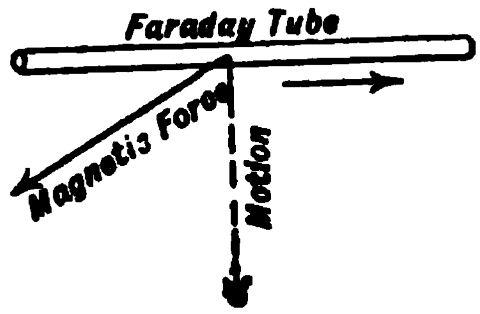


FIG. 308b.

bodies A and B with positive and negative charges respectively. These have electric lines or tubes connecting the charges. If we now connect A and B by a wire C, the ends of the lines will slide along the conducting wire, until the lines shrink to molecular lengths, when the charges combine. But as these lines shrink until their ends come together, there are magnetic lines *at right angles to the lines and to the direction of their motion* (Fig. 308*b*). In the case of a continuous electric current, there is a continual renewal of the electric lines, so that there is a stream of electric lines closing in along the conductor. According to this view, which is in agreement with the ideas of Faraday and Maxwell, the magnetic field is due to these moving electric lines.

433. Magnetic Effect of a Moving Electrified Body.—Rowland made in 1875 a fundamental experiment which showed that a charged body when moving at a high speed is equivalent in its magnetic effects to an electric current. His method was to charge a gilded vulcanite disk and spin it very rapidly. This produced a deflection of a sensitive magnetic needle. Measurements have shown that a moving charged body produces a magnetic field which is equal to the field produced, per unit of its length, by a linear conductor carrying a current eu, where e is the charge and u its speed. An electric discharge in a vacuum tube consists of streams of charged particles called cathode and canal rays, and it is found that these act in accordance with Rowland's experiment, that is, they are equivalent to electric currents, and are bent and deflected by a magnet like flexible currents. (See §552 on Conduction of Electricity in Gases.)

434. Electron Theory of Conduction.—Since moving charges have the same magnetic effects as electric currents, it is a natural supposition that a current consists essentially of a stream of charged particles, the combined magnetic effect of which constitutes the magnetic field associated with the current. In the discussion of views as to the nature of electric charges (§394) we have seen that the most probable hypothesis is that they consist of electrons or units of electricity which can be transferred from one body to another, an excess above the normal constituting a negative charge and a deficiency a positive charge. This hypothesis has been extended to the explanation of electric currents and it has been found to account fairly well for the facts. Since the electrons are very much smaller than the atoms of conductors, it would seem probable that the current must consist in a flow of electrons. If so, the flow must be from low to high potentials. In fact, high and low potentials have been defined by the work done in moving a charge of positive electricity. If it had been the unit of negative electricity that was referred to in the definition, the terms high and low, as applied to the potential of bodies, would have been reversed.

It is believed that in a metallic conductor many electrons are so entirely "free" or so loosely connected to atoms that they are easily set in motion by electric forces, whereas the much larger atoms, each of which remains positively charged when deprived of its normal number of electrons, move much more slowly. There is evidence that the electrons are moving in random directions with very high velocities when there is no current in the metal, for, under certain conditions, they can be ejected from the surface of the metal and something can be learned as to their velocities (§565). When a difference of potential is applied to the ends of the conductor, a drift of the electrons is superposed on their random motion and this drift constitutes the electric current. The drift or stream of electrons does not,

however, attain any very great velocity, since collisions between electrons and atoms are continually taking place.

From the above we can readily obtain an expression for the magnitude of an electric current. The moving electrons in any part of a circuit must produce the magnetic field associated with that part of the circuit. Now consider a wire of cross-section, a, and let the number of electrons per unit of volume be N, the charge of each being e in electromagnetic units. Then the total charge of these is Nae. Hence, by the result of Rowland's experiment, if the mean velocity in the direction of the stream is u,

$$i = Naeu$$

This expression has been tested in various ways, some of which will be referred to later.

MEASUREMENT OF CURRENTS

435. Galvanometers.—An instrument for measuring an electric current by its magnetic effects is called a galvanometer. If the instrument is only for detecting the presence of a current, it should in strict language be called a galvanoscope, but such an instrument is often called a galvanometer, or perhaps a detector galvanometer. There are two types of galvanometers in common use (*a*) the galvanometer with a movable magnetic needle and a fixed coil, and (*b*) the galvanometer with a movable coil and a fixed magnet. This last type is called the d'Arsonval galvanometer. Electrodynamometers, which are current measuring instruments depending on the magnetic action between two coils, one fixed and the other movable, form, strictly speaking, another type of galvanometers (see §531). The term ammeter or ampere-meter is used for special forms of graduated galvanometers. One of these will be described later (§440).

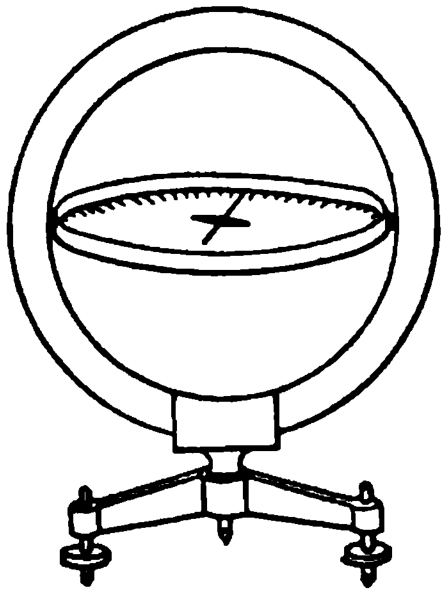


Fig. 309.

436. Tangent Galvanometers.—A tangent galvanometer consists of a circular coil, which is mounted with its plane vertical and set in the magnetic meridian, and a small magnetic needle, suspended horizontally at the center of the coil. The needle is in a compass box with a pointer and graduated circle so that its deflections can be read. The deflections are often measured by attaching a small mirror to the needle and observing the deflection of a beam of light on a scale. When an electric current passes in the

coil, the needle is under the action of the magnetic field of the earth, which is parallel to the coil, and of the magnetic field due to the current, which is at right angles to the coil. It takes a resultant position and makes an angle θ with the magnetic meridian. There are then two couples acting on the needle. The couple tending to turn it back into the magnetic meridian is $HM \sin \theta$, where H is the horizontal intensity of the earth's magnetic field and M is the magnetic moment of the needle (§376). The couple acting to turn the needle into the direction of the field of the coil is $\frac{2\pi n i}{r} M \cos \theta$, where $\frac{2\pi n i}{r}$ is the intensity of the field due to the coil (§428). When the needle is at rest, the two couples are equal, or

$$\frac{2\pi n i}{r} M \cos \theta = HM \sin \theta$$

Hence,
$$i = \frac{Hr}{2\pi n} \tan \theta$$

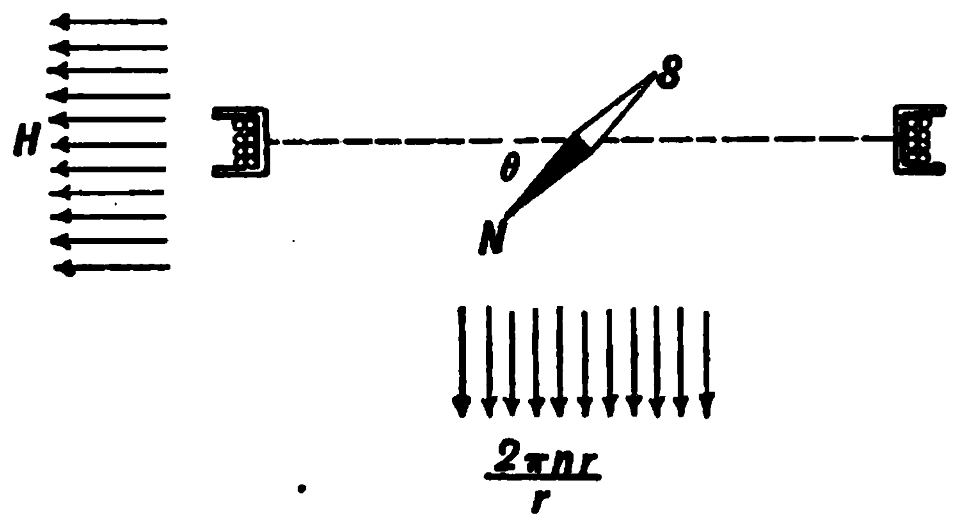


FIG. 310.

In this formula, the term $\frac{2\pi n}{r}$ depends only on the dimensions of the galvanometer and is represented by G, called the galvanometer constant. The formula then becomes

$$i = H/G \tan \theta = \mathrm{A} \tan \theta$$

The current is thus proportional to the tangent of the angle of deflection. If the current is to be measured in amperes, instead of c.g.s. electromagnetic units,

$$C \text{ (amperes)} = \frac{10Hr}{2\pi n} \tan \theta$$

In the above it has been assumed that the magnetic field due to the current is uniform for the region of the needle and equal to

the magnetic field calculated for the center of the coil. This is approximately true when the diameter of the coil is large compared to the length of the needle. It is usual to have a coil 25 cm. or more in diameter, and a needle a centimeter or less in length.

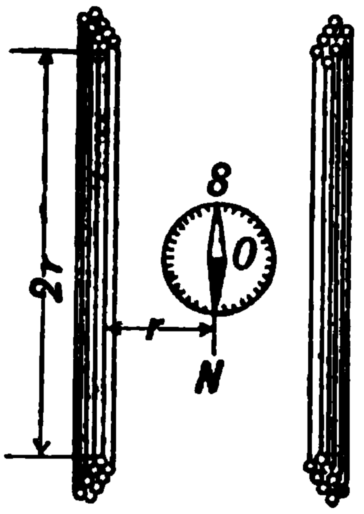


FIG. 311.

In the Helmholtz-Gaugain tangent galvanometer there are two equal vertical coils placed at a distance of the radius apart and with the needle on the axis midway between the two coils (Fig. 311). If the coils have more than a single turn they are generally wound on parts of cones which have their vertices at the midpoint of the coils. It can be shown that this arrangement gives a very uniform magnetic field for the region immediately around the midpoint.

Tangent galvanometers are used (1) to compare currents by comparing the tangents of the angles of deflection which they produce, and (2) to measure electric currents in absolute units. In the last case, the values of G and H have to be determined. To get G is a matter of simple measurement and arithmetic; the method of determining H has been given in §383.

437. Sensitive Galvanometers (Movable Needle Type).—A tangent galvanometer is primarily a standard instrument. Since its coil must be large to give the required uniformity of field at the center, very small currents will not produce readable deflections of the needle. To detect and measure small currents sensitive galvanometers have been devised in which (*a*) the action of the current on the needle is increased and (*b*) the directive action of the external field on the needle is weakened.

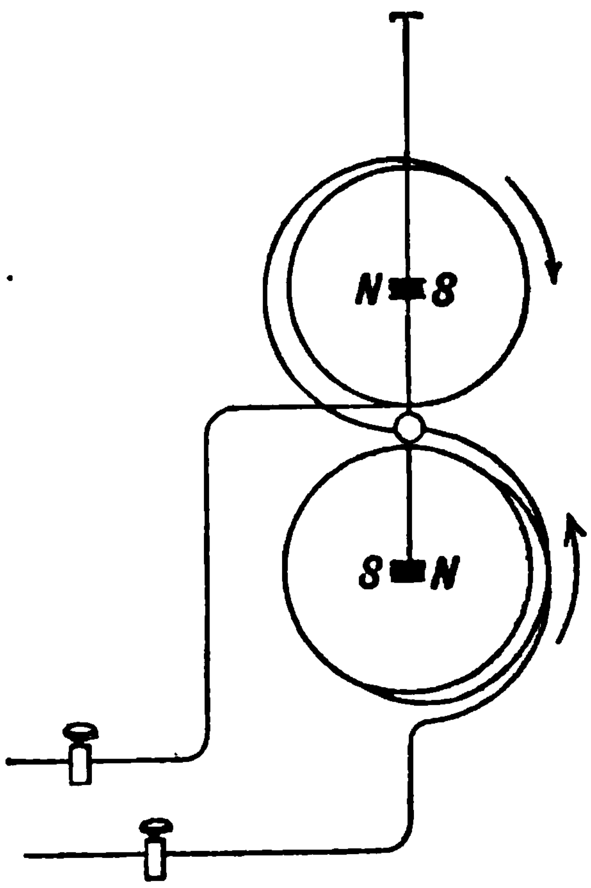


FIG. 312.

Among the most sensitive of galvanometers is the astatic mirror galvanometer of Professor William Thomson, Lord Kelvin, originally invented for receiving the weak signal currents of the Atlantic cable. The magnetic system consists of two magnetic

needles fixed parallel to each other on the same staff, but with the poles of the two needles oppositely directed (Fig. 312). The directive action of the field on the needle system is thus proportional to the difference in the magnetic moments of the two needles. In this galvanometer there are two coils, one surrounding the upper needle and one surrounding the lower needle of the astatic needle system. Each coil is double, and the needle hangs between the two parts of the coil. The coils are usually wound with very fine silk-covered copper wire so as to get the maximum length of wire on the coils near the needles. This makes coils of high electrical resistance, and hence such galvanometers are sometimes called "high-resistance" galvanometers. The suggested term "long coil" galvanometer, however, better describes

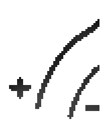


FIG. 313a.

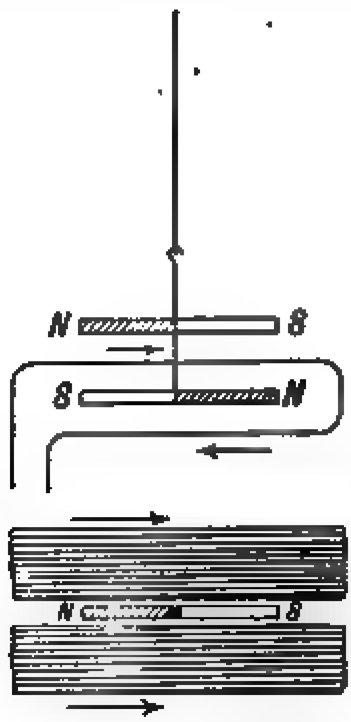


FIG. 313b.

these instruments. The exterior magnetic field is controlled in direction and in strength by one or more controlling magnets, placed on top of the instrument. By this means, the exterior magnetic field can be reduced to any extent. A small mirror is attached to the needle system and the deflections are read by a telescope and scale, or by a lamp and scale. For small deflections the currents are proportional to the deflections. A common sensitiveness of a galvanometer of this type is a deflection of 1 millimeter of a beam of light on a scale at a distance of 1

meter for 10^{-9} amperes though galvanometers of this kind are made with a sensitiveness of even 10^{-12} amperes. Another method of denoting the sensitiveness of a galvanometer is to give the resistance in the circuit when an e.m.f. of one volt causes a deflection of 1 mm. on a scale at 1 meter distance. Thus a galvanometer would be described as "sensitive to 1000 megohms."

One of the most sensitive of recent galvanometers is the Broca galvanometer. The needle system and coils in this galvanometer are indicated in Fig. 313*a*. The vertical magnets have consequent poles and may be very strong and are perfectly astatic if parallel. This produces greater sensitiveness and greater freedom from external disturbances.

In the above sensitive galvanometers, some damping device is necessary, in order to bring the needle to rest. In most cases air damping is used, a mica vane being for this purpose attached to the needle system. Magnetic disturbances due to commercial electric currents and machinery seriously limit the use of sensitive astatic galvanometers; indeed in many places they cannot be used unless magnetically shielded. This shielding is effected by means of a series of hemispheres or of cylinders of soft iron. (See §492.) Such instruments are called "iron-clad" galvanometers.

The simple "astatic needle multiplier" galvanoscope shown in sections in Fig. 313*b* was much used by earlier investigators. It can be made sensitive but is of course affected by external magnetic fields.

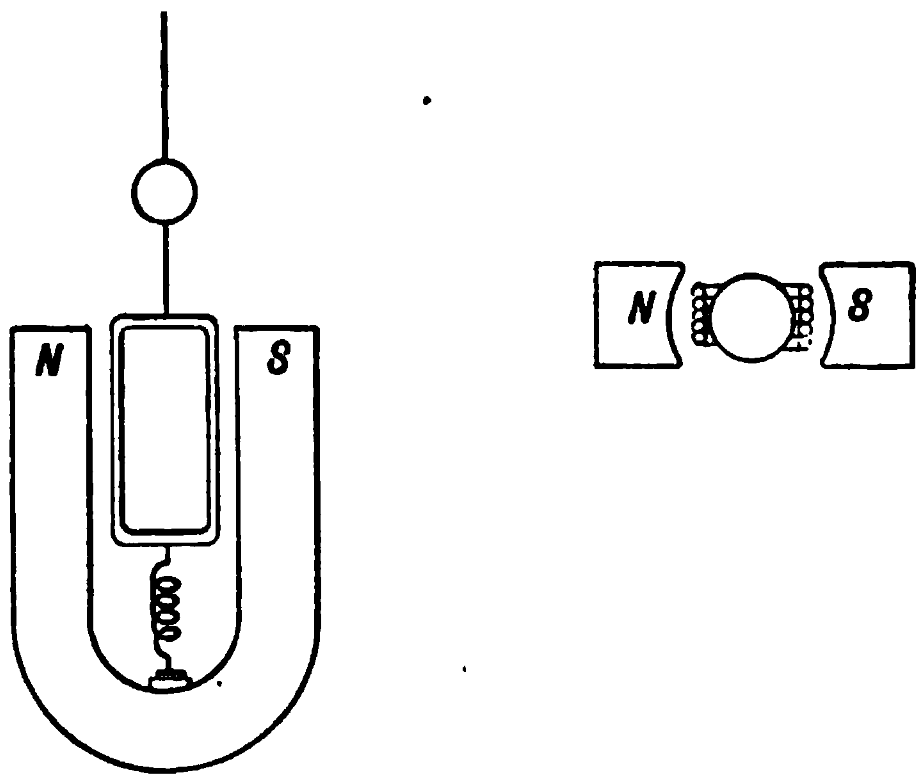


FIG. 314.

438. Moving-Coil or D'Arsonval Galvanometers.—A galvanometer of this type consists of a small coil suspended between the poles of a strong magnet by a phosphor-bronze or steel strip (Fig. 314). The upper connection for the current is

by the suspension strip, and the lower connection is by a loose loop of fine copper wire. The "controlling" force on the coil is the torsion of the suspension strip. When there is no current passing through the galvanometer, the plane of the coil is parallel to the magnetic field. When a current passes through the coil, one face becomes a north magnetic pole and the other face a south pole (§427). A current accordingly causes a deflection of the coil in the magnetic field. In the more sensitive instruments these deflections are read by one of the mirror and scale methods. The great advantage of the d'Arsonval galvanometer is that it is practically free from external magnetic disturbances. The quick damping of the vibrating coil is also a great convenience. This damping is due to the reaction of the current induced in the moving coil itself on closed circuit, or in a closed metallic frame inserted inside the coil. (See Lenz's law §501.) D'Arsonval galvanometers are so much more convenient to use than astatic galvanometers, that they have almost completely superseded the later instruments for most electrical measurements. A common sensitiveness for a d'Arsonval galvanometer is 10^{-8} amperes for 1 mm. deflection at a meter distance, though in special instruments a sensitiveness of 10^{-10} amperes is reached.

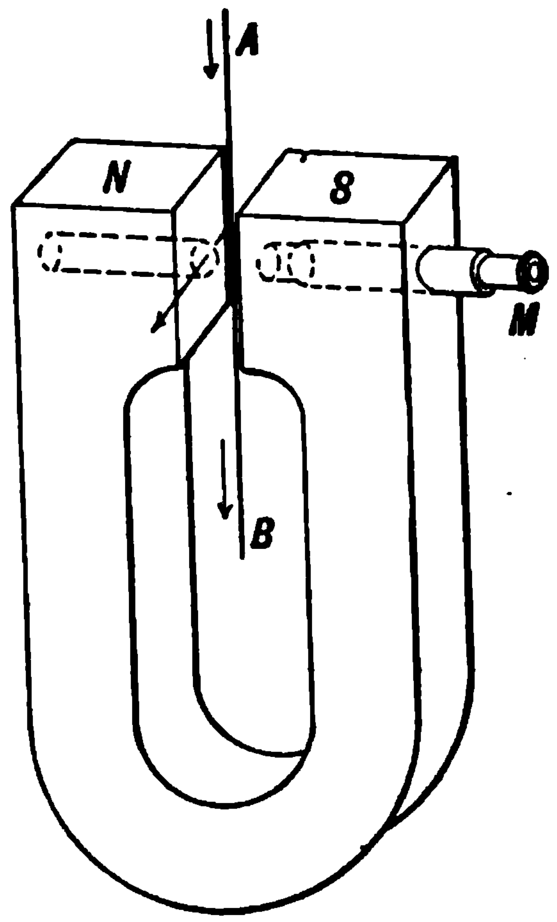


FIG. 315.

In the Einthoven "thread" galvanometer, there is a single fine wire stretched across the field between the poles of a strong magnet (Fig. 315). A current through this wire causes a deflection of the wire, owing to the mutual force between a current and a magnet (§528). This deflection is read by a microscope with a micrometer eyepiece. The magnet may be a permanent magnet or an electromagnet. Instead of a metallic wire, a silvered quartz fiber is used in very sensitive Einthoven galvanometers. A current as small as 10^{-12} amperes can be detected by a galvanometer of this type.

439. Ballistic Galvanometers.—A ballistic galvanometer is a sensitive galvanometer which has (1) a long period of swing and (2) little air friction or other damping action on the needle or moving coil. Either type of galvanometer may be used for ballistic purposes. The period suitable for most cases is from 6 to 10 seconds for a single oscillation. The long period is obtained by loading the needle or coil, thus increasing its moment of inertia. The instrument is used to measure the total quantity of electricity in a transient current, such as we get in the discharge of a condenser or in an induced current of short duration (§505). The principle is that the transient current produces an impulse (§87) which is proportional to it or Q; and further that this impulse is proportional to the first throw θ of the needle, *provided the needle does not move appreciably during the time of the discharge.* From this we get $Q = k\theta$, or the total electric quantity of this discharge is proportional to the first throw of the needle or moving coil. The complete theory of the ballistic galvanometer with discussion of the conditions of its use, is given in the larger manuals.

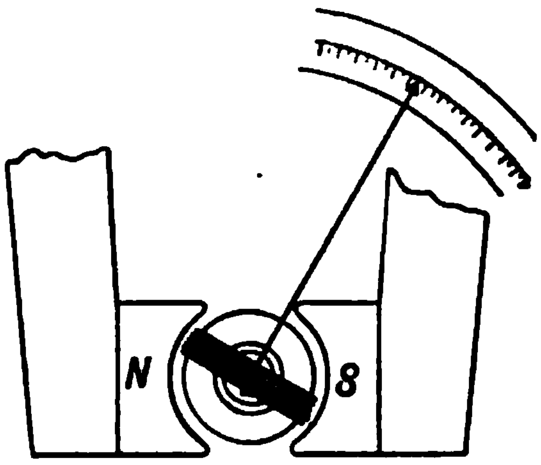


FIG. 316.

440. Weston Ammeter.—This is a D'Arsonval galvanometer with a coil mounted on steel points in agate bearings and controlled by a flat spiral spring (Fig. 316). The deflections are read by the movement of a light pointer that moves over a scale which is graduated directly in amperes. The coil will carry only very small currents, and so shunts (§452) are used for the larger currents. Thus only a known fraction of the total current passes through the needle system.

ELECTROMOTIVE FORCE AND RESISTANCE

441. Units of Potential Difference and Electromotive Force.—When a quantity of electricity, q, passes from a point at the potential V_1 to a point at the potential V_2, it does an amount of work expressed by $W = q(V_1 - V_2)$ (§402). When there is a current in a conductor a quantity of electricity, $q = it$, flows in time t through every cross-section of the conductor, and, as regards the part of the conductor between points at potential V_1 and V_2, the effect is the same as if the quantity q had passed from one

point to the other. Hence the work in this part of the conductor is $W=it(V_1-V_2)$ and therefore $V_1-V_2=W/it$. If W is expressed in ergs, t in seconds, and i in electromagnetic units of current, V_1-V_2 is in electromagnetic units of difference of potential.

The measure of the electromotive force in a closed circuit has been already defined (§423) as the work done in taking unit quantity of electricity once around the circuit. In time t the quantity $q=it$ passes through every cross-section of the circuit and the effect is the same as if q had passed once around the circuit. Hence $W=Eit$ or $E=W/it$. Hence a difference of potential and an electromotive force are quantities of the same kind and must be expressed in the same unit. The latter term is the more general, since it applies to a complete circuit, and the electromotive force in a circuit in which the only generator is a voltaic cell is equal to the sum of the potential differences at the contacts (§425). The difference of potential between two points of a conductor which contains no generator is frequently called the electromotive force acting in that part of the conductor or simply the electromotive force between the points.

In accordance with the equation $E=W/it$, stated above, we define the unit of electromotive force as follows:

The electromagnetic unit e.m.f. exists between two points when one erg of work is done by one electromagnetic unit of current flowing for one second between the two points.

Experiment shows that the electromagnetic unit of e.m.f. or d.p. is about $\frac{1}{3}\times 10^{-10}$ the electrostatic unit d.p. already defined (§405). As a practical unit of e.m.f. we use the *volt.* The volt is 10^8 times the c.g.s. e.m. unit of e.m.f. It follows from the above that the product tiE is in ergs, when i and E are expressed in electromagnetic units, and t in seconds. An ampere (10^{-1} e.m. units) flowing between two points with an e.m.f. of one volt (10^8 e.m. units) will then do 10^7 ergs of work per second, or 1 joule per second (§55). Thus i (in amperes) $\times$ E (in volts) $=$ W (in joules per second) $=$ W (watts). This work is transformed into heat (§458), or chemical energy (§462) or into mechanical energy (§§534, 536).

442. Conductivity and Resistance. Ohm's Law.—Experiments show that in a circuit with a continuous and steady current, *the*

electric current is directly proportional to the electromotive force, the constant of proportionality depending only upon the materials and dimensions of the circuit. This fact is stated by the equation $i = CE$, where i is the current, E the e.m.f., and the constant of proportionality C is called the *conductance* of the circuit. We more commonly use a different constant R, the reciprocal of the conductance C, and write the equation in the form $i = E/R$. R is then defined as the *resistance* of the circuit. The above is called "Ohm's law," and was first formally stated by G. S. Ohm in 1828. If we write the equation in the form $R = E/i$, we get the important statement: *the electrical resistance of a conducting circuit is the constant ratio between the e.m.f. and the current in the circuit.* The resistance of a circuit therefore does not depend upon the size of the current; that is, so long as the dimensions and physical properties of the circuit remain unchanged, the resistance is a constant for the circuit.

Ohm's law holds not only for the whole circuit, but also for any part. Thus if a wire AB, which forms part of a circuit has a current i in it, and an e.m.f. or difference of potential of E_{AB} between its ends, its electrical resistance is $R_{AB} = E_{AB}/i$. From this relation it follows that R is unity when E and i are both unity. That is, *the electromagnetic unit of resistance is the resistance of a conductor in which one electromagnetic unit of current is produced by an electromagnetic unit difference of potential*, or e.m. unit e.m.f.

The practical unit of resistance is the *ohm*. The ohm is *the resistance of a conductor in which a current of one ampere is produced by a difference of potential of one volt.* This is the "absolute" ohm as distinguished from the legal or international ohm (§449). We thus see that $R\ (ohms) = \dfrac{E\ (volts)}{i\ (amperes)}$. Since the volt $= 10^8$ c.g.s. e.m. units and the ampere $= 10^{-1}$ c.g.s. e.m. units, the ohm must equal 10^9 c.g.s. e.m. units of resistance.

443. Extension of Ohm's Law.—It is sometimes necessary to calculate the current in a part of a circuit when the part in question contains a voltaic cell (or some other form of generator). For this purpose we must use a more general form of Ohm's law which can be found as follows. Consider a circuit $ACBD$ which contains a cell, D, the e.m.f. of which is E. Applying Ohm's law to the whole circuit and to the part BCA, in which there is no

generator, and denoting the resistance of the part ADB by R_{ADB} and that of BCA by R_{BCA}, we get

$$E=i(R_{ADB}+R_{BCA})$$
$$V_B-V_A=iR_{BCA}$$

From these we get by subtraction

$$E+V_A-V_B=iR_{ADB}$$

or

$$i=\frac{E+V_A-V_B}{R_{ADB}}$$

Thus to get the current that flows from A to B we must add to the excess (positive or negative) of the potential of A over B the e.m.f. of the generator D and divide by the total resistance of the part ADB, including that of the generator.

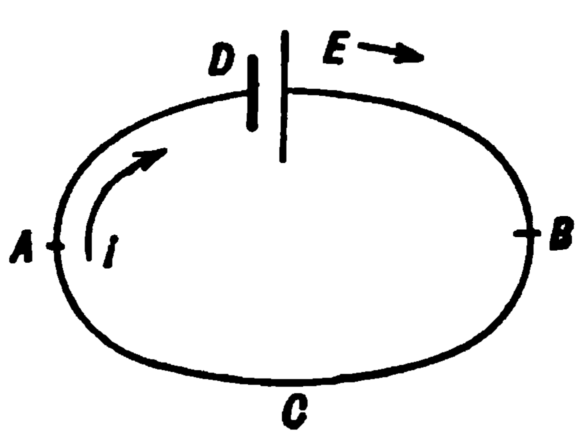


FIG. 317.

The above shows that, in applying Ohm's law to a part of a circuit, the electromotive force in that part cannot be taken as the difference of potential at the ends when the part contains a generator. We shall see later that there are cases in which every part of a circuit must be regarded as a generator (§§499, 500).

444. Specific Resistance. Conductivity.—The resistance of a conductor varies directly as its length L, inversely as its cross-section A, and directly as a quantity ρ, called the specific resistance or the "resistivity" of the material, that is $R=\rho\frac{L}{A}$. *The specific resistance of a substance is the resistance of a bar of the substance one centimeter long and of one square centimeter cross-section.* The table on p. 394 gives the specific resistances at 0°C. of a number of materials ordinarily used in the arts. From this table and the length and cross-section of a wire we can calculate its resistance by the above formula.

The specific resistance of a substance varies with its temperature, and, in the case of solids, with properties which depend upon previous treatments, such as hardness, temper, structure, etc. The effect of temperature will be discussed in a later section (§447). The effect of previous treatment cannot be stated in a simple form. Hence different samples of the same substance may show quite different specific resistances.

The specific electrical conductivity of a substance is the recip-

rocal of its specific resistance. Thus taking the specific resistance of certain copper as 1.59×10^{-6}, its specific conductivity is $1/\rho = 6.29 \times 10^{5}$.

SPECIFIC RESISTANCES

Material.	Resistance in ohms and cms.	Temperature coefficient per C°.
Aluminum	3×10^{-6}	.0043
Copper	1.5×10^{-6}	.0040
German silver	20×10^{-6}	.0004
Iron	10.5×10^{-6}	.0062
Manganin	42×10^{-6}	.00002
Mercury	94×10^{-6}	.00075
Platinum	8.9×10^{-6}	.00366
Silver	1.5×10^{-6}	.00377

445. Weight of Wire.—Calculations of wire tables are also conveniently made on the basis of the length and the weight of the wire, the weight of course being directly proportional to the cross-section of the wire. The reference unit in this case is the wire which is one meter long and weighs one gram, called the "meter-gram." The Bureau of Standards has determined from many experiments on *commercial* copper, that the "standard" meter-gram of annealed copper wire at 20°C. has a resistance of 0.153022 ohms. This corresponds to a density for copper at 20°C. of 8.89, and a specific resistance of 1.72128 $(10)^{-6}$ in ohms and centimeters or a specific conductivity of 5.8096 $(10)^{-4}$ c.g.s. e.m. units at 20°C. The mean temperature coefficient at 20°C. is taken as $\alpha = 0.00383$. The conductivity of hard-drawn copper wire is about 2.7 per cent. less than that of annealed copper wire. The use of weights instead of sectional area is to be recommended, because the weight can be determined with greater accuracy, particularly in the case of fine wires and in the case of wires of irregular shapes of cross-sections.

446. Resistance of Alloys.—The resistance of a substance is in general increased by even small amounts of foreign substances. Thus it has been found that one-half per cent. of carbon changes the conductivity of copper by 20 per cent. Many experimenters have studied the resistances of mixtures of metals, known as alloys, both on account of the theoretical interest and the practical importance of the results. For one group of metals, lead, tin, cadmium and zinc, it is possible to calculate the resistance of the alloy as the mean resistance of the volume constituents. In the case of alloys of practically all other metals, the resistance of the alloy is considerably greater than that of any of its constituents. Another most

important property of certain alloys is that the change of resistance with temperature is very small. These properties make such alloys as constantan, manganin, platinoid, etc., very valuable for resistance standards and for rheostats.

447. Resistance and Temperature.—The electrical resistance of pure metals increases as the temperature rises. The increase per degree from 0° to 100°C. is a certain fraction of the resistance at 0°C. This fraction is called the *temperature coefficient of resistance.* The above law may also be stated in the form of the equation, $R_t = R_o\ (1 + \alpha t)$. This when plotted gives a straight line.

For larger ranges of temperature a formula involving a second constant β, and of the form

$$R_t = R_o(1 + \alpha t + \beta t^2)$$

must be used.

448. Resistance Thermometers.—The resistance of a coil of wire, being a function of the temperature, can be used to determine temperatures. Platinum has been found to be the best metal for resistance thermometry. The advantages of a platinum resistance thermometer are sensitiveness, and the wide range of temperature that can be measured (from the lowest temperature to +1200°C.). A platinum thermometer can also be of almost any size and shape, and the thermometer coil can be at a distance from the resistance bridge and the observer. Fig. 177 shows a standard resistance thermometer as devised by Callendar.

449. Resistance Standards.—It follows from the definition of the ohm that the "absolute" measurement of the resistance of a conductor consists in determining the ratio of the e.m.f. (volts) and the corresponding current (amperes) in the conductor. To make such a measurement with high accuracy is not a simple process. But to get the ratio of two resistances is, as we shall see later (§456), a relatively simple measurement and one that can be made easily with very high accuracy. Hence the ordinary process of determining the resistance of a conductor is one of comparing its resistance with a "standard resistance."

Standard resistances are of two classes, (1) the prime standard. a mercury resistance and (2) secondary standards in the

form of coils of wire either (*a*) single coils, or (*b*) groups of coils mounted in boxes or cases, and hence called *resistance boxes.*

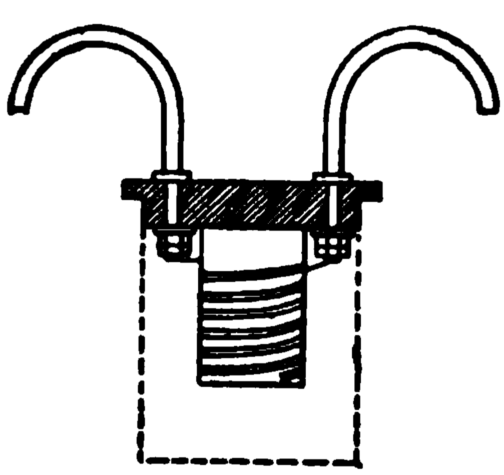


Fig. 318.

The prime standard is defined so that it can be reproduced from the specifications of materials and dimensions only. At an International Congress of Electricians held at Chicago in 1893, in which all civilized nations were represented, it was recommended that "*the international ohm be the resistance offered to an unvarying electric current by a column of mercury at the temperature of melting ice,* 14.4521 *grams in mass, of a constant cross-sectional area and of the length of* 106.3 *centimeters.*" The cross-sectional area of such a column of mercury is 1 square millimeter. This has been adopted by all nations as the *legal ohm.* The ohm as thus defined by law was as near the absolute ohm as measurements could fix it at the time.

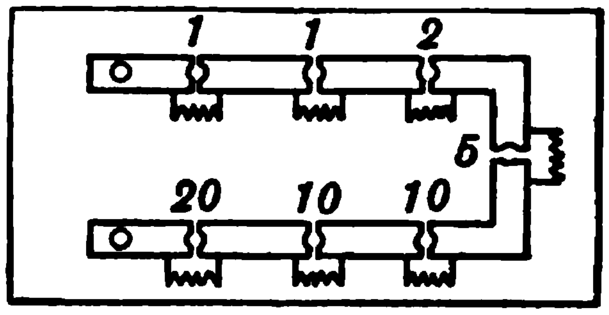


Fig. 319.

Resistances in the form of wire coils are the most convenient working standards. First we have single coils made in a form shown in Fig. 318. They are made so that they can be immersed in an oil bath of constant temperature, and are provided with large copper terminals to dip in mercury cups. Resistances of this kind are used primarily for calibrating the working resistance boxes. They should be supplied with certificates of calibration from one of the national calibrating laboratories, such as the U. S. Bureau of Standards, The Reichsanstalt of Germany, or the National Physical Laboratory of Great Britain.

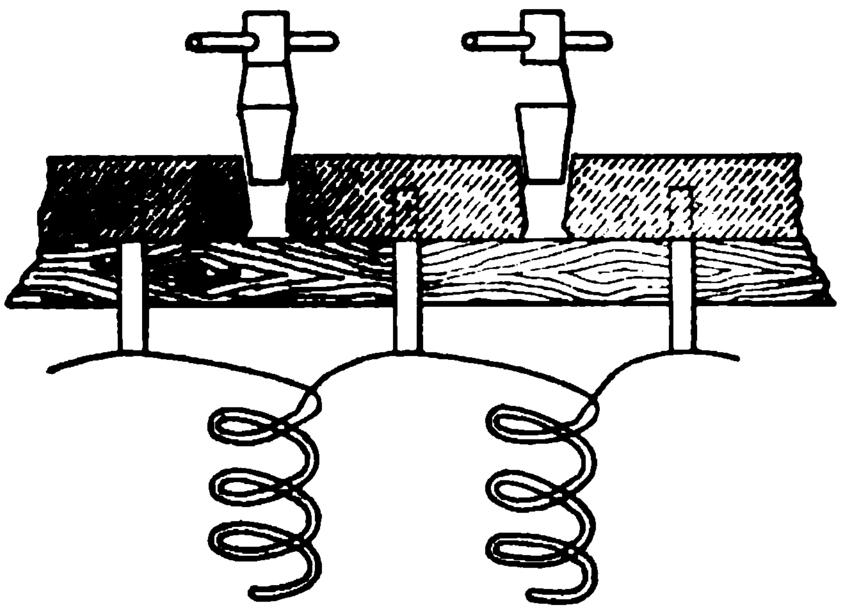


Fig. 320.

For general laboratory purposes resistance coils are mounted in boxes as shown in Fig. 319. On the ebonite top there are a

series of heavy brass blocks and the ends of the coils are joined to these blocks, so that the current entering at one terminal passes from block to block through each resistance coil in turn. Any coil can be cut out of the circuit by bridging the brass blocks with a metal plug (Fig. 320). Instead of plugs, a lever with sliding contacts is used successfully in some recent resistance boxes. Most of the high-grade resistance boxes are now wound with manganin wire. A resistance coil is always wound inductionless (Fig. 321), that is, the coil is wound back on itself so as to avoid magnetic effects and self-induction (§508).

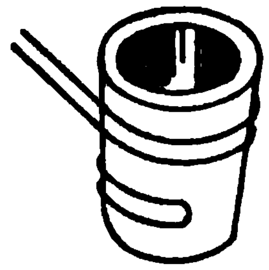


Fig. 321.

450. Resistance of Combinations of Conductors. (a) *Series Arrangement.*—The total resistance of a number of conductors connected "in series" is equal to the sum of the resistances of the individual conductors, that is, $R=r_1+r_2+\cdots$, where R is the total resistance, and r_1, r_2, r_3, etc., are the resistances of the individual conductors. For the differences of potentials between the ends of the individual conductors are ir_1, ir_2, ir_3, etc., and the total difference of potential is iR. Hence $iR=i(r_1+r_2+\cdots)$ and $R=r_1+r_2+\cdots$.

(b) *Parallel Arrangement.*—When a number of conductors connect the same two points, the resistance of the combination of conductors is given by the expression $1/R=1/r_1+1/r_2+\cdots$, where R is the resultant resistance and r_1, r_2, r_3, etc., are the resistances of the individual conductors.

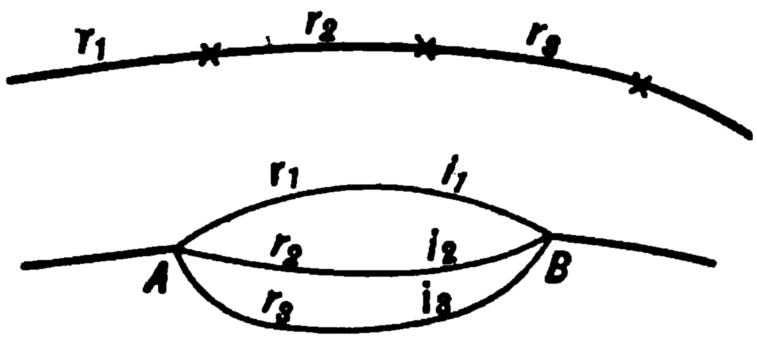


Fig. 322.

Let E be the electromotive force between the two points A, B (Fig. 322) and R, the resultant resistance of the separate resistances r_1, r_2, r_3, etc., in parallel. Then the currents in the separate branches are

$$i_1=E/r_1,\ i_2=E/r_2,\ i_3=E/r_3, \text{ etc.}$$

or the total current is

$$I=i_1+i_2+\cdots=E(1/r_1+1/r_2+\cdots)=E/R$$

or,

$$1/R=1/r_1+1/r_2+1/r_3+\cdots.$$

We thus get the statement, *the sum of the reciprocals of the separate resistances in parallel is equal to the reciprocal of the resultant resistance.*

In the above proofs we have assumed, in addition to Ohm's law, that the algebraic sum of the currents flowing toward a point such as A is zero, that is, considering outward-flowing currents as negative, $+I-i_1-i_2-i_3=0$.

451. Kirchhoff's Laws.—The laws for *steady* currents in branched circuits, one of which has been assumed above, have been stated by Kirchhoff in the following general form: (1) The algebraic sum of the currents which meet at a point is zero, or $\Sigma i=0$. (2) In any closed circuit, the algebraic sum of the products of the current and resistance in each of the conductors in the circuit is equal to the electromotive force in the circuit, $r_1i_1+r_2i_2+r_3i_3=E$.

The first law is equivalent to the statement that when the currents in a network are steady there is no accumulation of electricity at any junction—all that flows in must flow out. The second law can be deduced

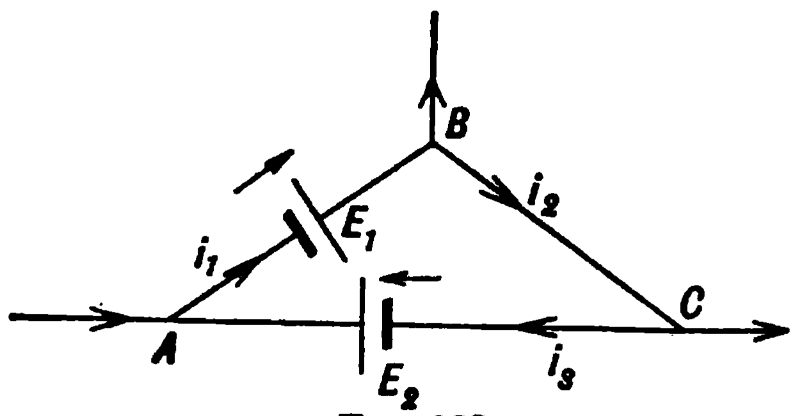


FIG. 323.

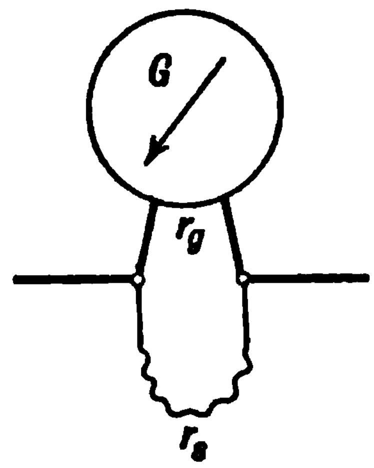


FIG. 324.

from the extended form of Ohm's law stated in § 443. For let ABC be any closed circuit in a complex network and let the currents resistances and e.m.f.'s in the branches be as indicated in Fig. 323. Then

$$\begin{aligned} i_1r_1 &= V_A - V_B + E_1 \\ i_2r_2 &= V_B - V_C \\ i_3r_3 &= V_C - V_A + E_2 \\ \hline \Sigma ir &= \Sigma E \end{aligned}$$

It is to be noted that, in applying the second law, one direction around the circuit must be chosen as positive, and each current and e.m.f. must be considered as positive or negative according as it is in this or the opposite direction respectively.

452. Branched Circuits, Shunts.—The principle of parallel circuits is taken advantage of in shunts for apparatus. Thus in the case of a galvanometer, it is often necessary to measure a current which is much larger than it is desirable to pass through the instrument. A branch circuit or *shunt* of known resistance r_s is put in parallel with the galvanometer (Fig.

324). The current is thus divided into i_s and i_g, where $i_s+i_g=i$, and $i_s r_s = i_g r_g$, whence it readily follows that $i_g/i = r_s/(r_g+r_s)$. Thus to have only 1/10 of the total current pass through the galvanometer, we use a shunt having 1/9 of the resistance of the galvanometer.

453. Milli-ammeters as Voltmeters.—A milli-ammeter with a high resistance in series is used as a voltmeter. The instrument is joined in parallel across the terminals of the generator or circuit for which the e.m.f. is to be determined (Fig. 325). The current through the milli-ammeter is proportional to the e.m.f. between its terminals, that is, $i_1 = E/R_1$. If R_1 is so large that introducing it as a branch circuit does not change the current in the main circuit appreciably, then the readings of the milli-ammeter are practically proportional to e.m.f. of the circuit for all currents, and the scale of the voltmeter can be graduated in volts. The resistance of the Weston voltmeter for 150 volts is about 15,000 ohms, and hence takes a current of 0.01 ampere or less. The change of potential caused by introducing this between the terminals of circuits of moderate resistances is for most purposes negligible.

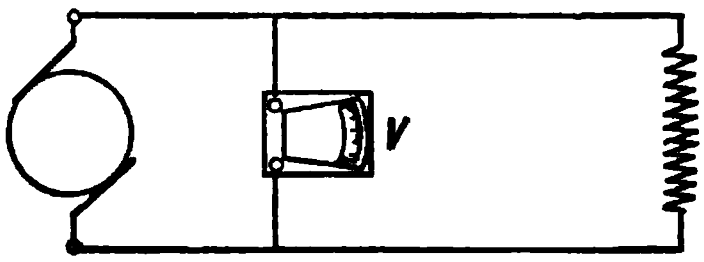


FIG. 325.

454. Fall of Potential in a Circuit.—When a current flows through a wire, there is a decrease or fall of potential in the direction of the current, for otherwise there would be no flow of electricity. Between any two points x and y of a conductor the fall of potential is $E_{xy} = iR_{xy}$, and hence we get the statements:

(a) With a constant current the fall of potential is proportional to the resistane between the two points.

(*b*) With a given resistance the fall of potential is proportional to the current between the two points.

The above simple deductions from Ohm's law are used continually in applied electricity. Thus with a given current to be transmitted from a machine to a distance, and with a certain allowable fall or "drop" of potential, the resistance (and hence the size) of the conducting wire can be directly calculated. Two of the most important instruments for electrical measurements, the potentiometer and the Wheatstone bridge, are based directly on the above laws. These instruments are described in the next sections.

455. The Potentiometer.—This instrument in its simplest form consists of a long uniform wire AB through which a constant current flows from a battery M (Fig. 326). There is a fall of

potential from A to B and, the wire being uniform, the fall of potential between two points is proportional to the length of wire or the resistance between the two points. If two points A and C on the wire be joined to a galvanometer G, there will be a current through AGC, as shown by the deflection of the galvanometer. If we now introduce an opposing e.m.f., E_x, (a galvanic cell, a thermo-element, etc.) in the galvanometer circuit, and find the point C, when there is no current in the galvanometer, we know that the fall of potential between A and C is equal to the e.m.f. E_x. In the same way we find a point D, such that the difference of potential between A and D is equal to the e.m.f., E_y, of a second galvanic cell.

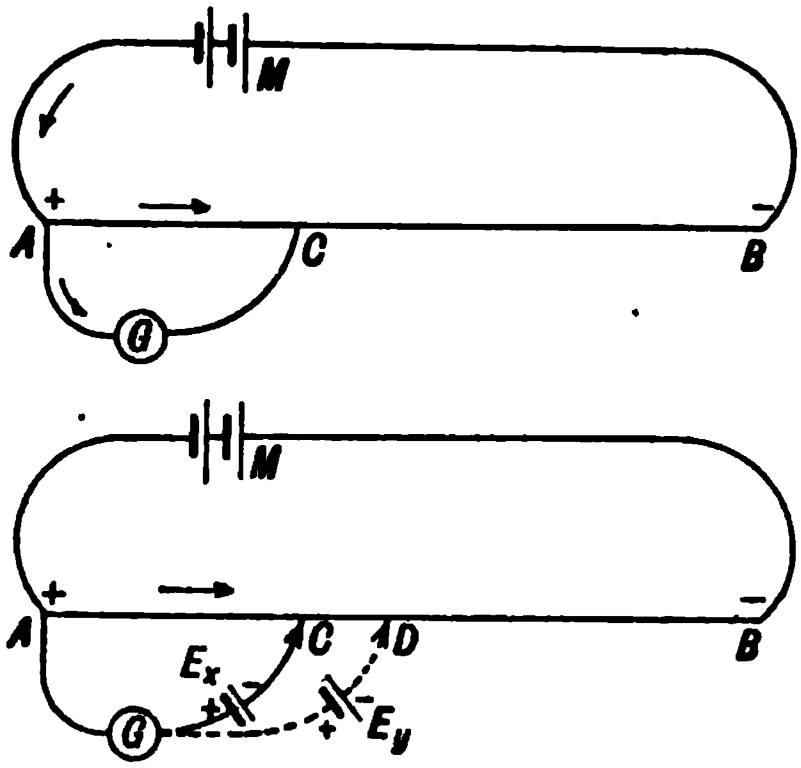


FIG. 326.

Hence $E_x : E_y$ :: resistance AC : resistance AD,
:: length AC : length AD

In this way two electromotive forces can be compared and by using a standard cell, such as a Clark or a Weston cell (§473) of known e.m.f., we can thus measure any other e.m.f. In potentiometers of the highest precision, the exposed wire is replaced by resistance coils in a box.

456. The Wheatstone Bridge.—This is an arrangement for getting a proportion between four resistances. At a point A (Fig. 327), the circuit divides into two branches, ACB and ADB. There is the same fall of potential, E_{AB}, along each branch. Hence we can find for any point C on the upper branch a corresponding point D on the lower branch, such that the potentials of C and D are the same. When two such points

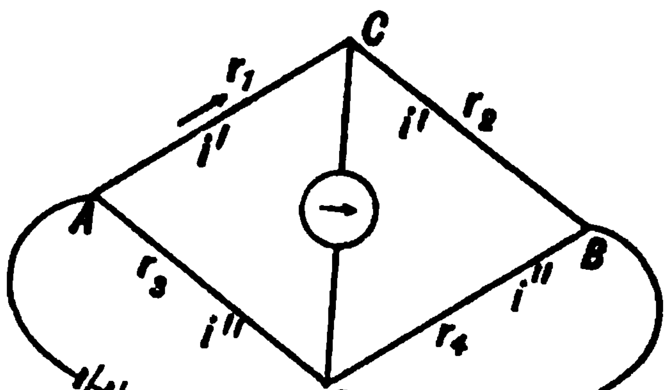


FIG. 327.

are joined through a galvanometer, the instrument shows no deflection. Let r_1, r_2, r_3 and r_4 be the resistances of the four parts AC, CD, AD, and DB. Then the current in r_1 and r_2 is i' and the current in r_3 and r_4 is i''. The falls of potential are $i'r_1$, $i'r_2$, $i''r$ and $i''r_4$. Since C and D are at the same potential,

$$i'r_1 = i''r_3$$

and

$$i'r_2 = i''r_4$$

By division we then get

$$r_1/r_2 = r_3/r_4$$

Thus

$$r_4 = \frac{r_2}{r_1} r_3$$

Hence, knowing the three resistances r_1, r_2 and r_3, we can get the fourth resistance, or, knowing the ratio r_2/r_1 and the resistance r_3, we can get the fourth resistance r_4.

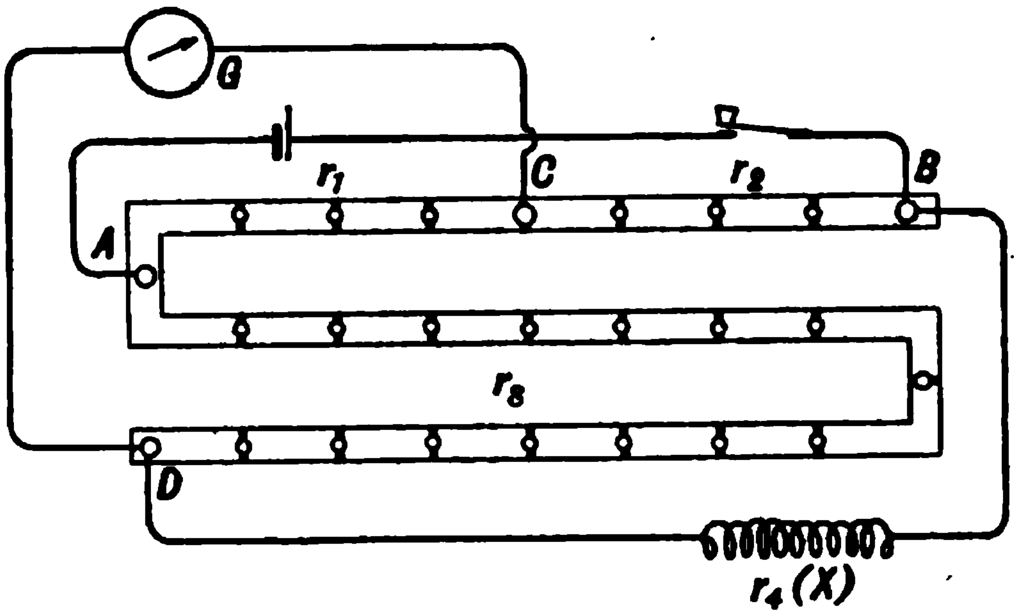


FIG. 328a.

Figs. 328*a* and 328*b* show forms of Wheatstone bridge; 328*a* a box bridge, often called a "post office" box bridge, 328*b* the "slide wire" or "meter" bridge. In this last form, the two resistances, r_1 and r_2 are the two parts AC and CB of the uniform wire AB. The ratio r_2/r_1 is thus given by the ratio of the lengths CB and AC.

It is evident that the galvanometer and the battery can be interchanged according to the above explanation of the Wheatstone bridge. The best arrangement for sensitiveness depends upon the relative resistances of the arms, galvanometer and battery. The rule which is proved in larger treatises, is that the most sensitive arrangement, when the galvanometer resistance is greater than the battery resistance, is gotten by placing the galvanometer between the junction of the two higher resistances and the junction of the two lower resistances.

457. Electron Explanation of Electric Resistance.—On the electron theory (§ 394) an electric current consists of a stream of electrons, each of which is performing more or less random motions but is, on the whole, moving forward in the direction of the electric force. We must, however, account for the fact that the stream moves forward at a steady rate and not, as might be expected, with an acceleration. The explanation is to be found in the frequent collisions between electrons and atoms. Between two successive collisions an electron has an acceleration in the direction of the electric force, but the forward speed is being continually checked by collisions, as when a man seeks to make his way rapidly through a crowd. Thus the forward motion is limited by the average forward velocity attained between collisions, and this is, of course, proportional to the electric force. Now we have already seen that i is proportional to Neu, where u is the average forward velocity of the stream. Hence, assuming that N is constant in a

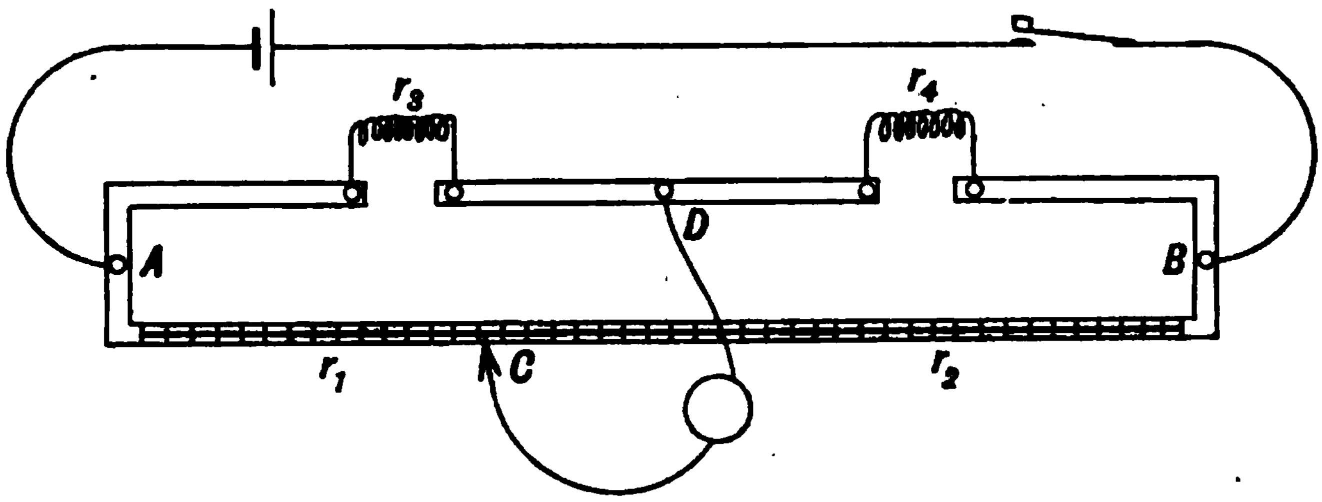


Fig. 328b.

conductor in a constant physical condition and that e is invariable, we see that the current is proportional to the electric force and this is Ohm's law.

In conductors of different materials the frequency of collisions between electrons and atoms must differ greatly, depending on the average distance between atoms. Hence under equal electric forces the values of u must also differ. We have thus a natural explanation of differences in conductivity and resistivity. The value of e does not vary and there is good evidence that N does not differ much in good conductors, such as metals, though it must be very different in very poor conductors.

HEATING BY ELECTRIC CURRENTS

458. Joule's Law.—That a current heats a wire through which it passes was observed early and in 1841 James Prescott Joule proved by experiments that the heat produced varied directly as the square of the current and directly as the resistance of the wire, or H is proportional to i^2R. It was shown later that if the heat H is expressed in calories, R in ohms, and i in amperes, then H (calories) $= .238Ri^2t$.

The above can be deduced directly from the energy relations involved. From the definitions of the units of electromotive

force and current it follows that the work in joules done by a current of i amperes, when it flows for t seconds between points the difference of potential of which is e volts, is $W = iet = i^2Rt$ by Ohm's law. Dividing this by the mechanical equivalent of heat (4.2 joules per calorie), (§290) we get H(cal.) $= W/4.2 = .238\ i^2Rt$. Where electrical energy is transformed into heat in connecting wires, it is ordinarily a loss and dissipation of energy, and so a wire of as low resistance as is economically profitable is used. Electric currents are, however, widely used to produce heat for important applications, such as in electric lighting (arc and incandescent lamps), in furnaces for metallurgical purposes, cooking, etc., in fuses of various kinds (safety, blasting, etc.), in hot-wire ammeters, etc., etc.

459. Incandescent Lamps.—The ordinary incandescent lamp consists of a high-resistance filament of carbon, tantalum, or tungsten, enclosed in an exhausted glass bulb, and arranged with terminals so that when the metallic base of the bulb is inserted in a socket connected to electric mains from a power station, a current flows through the filament. The current heats the filament to incandescence and the filament thus becomes a luminous source. The efficiency of an incandescent lamp, that is, the percentage of electrical energy transformed into visible luminous energy, is not high at the best, but is increased by raising the temperature of the filament. Increased efficiency thus becomes largely a question of finding filaments that will stand high temperatures. Tungsten lamps require a little over a watt of electrical power per candle power.

In the Nernst lamp the filament is a rod or "glower" made of refractory earths (oxides of zirconium and yttrium) and is a conductor only when heated. The "glower" is heated by an auxiliary "heater" until it becomes a conductor, and it is then maintained at incandescence by the current. No exhausted bulb is required for this lamp since the materials of the glower do not oxidize in the air.

460. The Electric Arc.—If two carbon rods AB and BC (Fig. 329) are in an electric circuit, and a current of several amperes passes across their contact point BC; it is found that the current continues when the carbons are separated, leaving a gap of a few millimeters between the ends B and C. A bluish "arc" is formed across the gap BC, and at the same time the ends of the carbons become incandescent. If the current is continuous, the positive

carbon takes a cup form known as a "crater" and is very much the hotter of the two carbons, and also gives off more light. The highest temperatures that have been produced artificially are those of the crater of the electric arc, estimated by Violle at 3500°C. The passage of the current is through carbon vapor formed between the carbons.

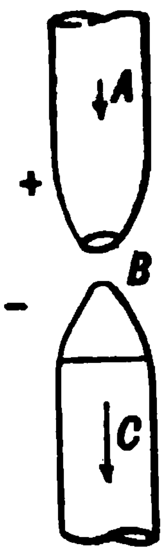


Fig. 329.

The complete discussion of the electric arc involves the theory of the conduction of electricity through gases. But two or three facts may be noted. The current is carried by ions, or charged particles, due to the dissociation of the vapors, and the ions move faster in one direction than in the other. If the anode (positive) is kept cold no arc will be formed, but an arc can be formed with a cold kathode (negative). The high temperature of the anode is probably due to the bombardment of the ions from the kathode. With direct or continuous currents an electromotive force of about 45 volts is necessary to maintain the arc satisfactorily. Currents from 6 to 50 amperes are used, the larger currents calling for larger carbons. Alternating currents may also be used, but in the case of the alternating current, both electrodes are alike, and the temperature of neither is as high as that of the anode with the equivalent continuous current.

The carbons are consumed by oxidation in the electric arc, the positive carbon wearing away about twice as rapidly as the negative carbon. The consumption of the carbons is greatly decreased by limiting the supply of air to the arc as in the so-called "enclosed arc lamps."

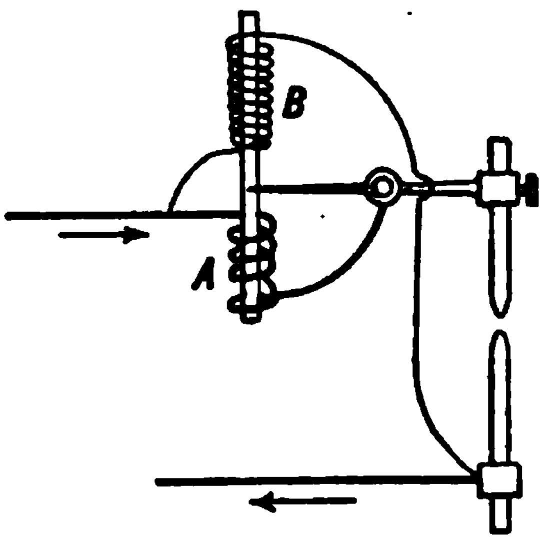


Fig. 330.

Arc lamps are supplied with a mechanism for automatically feeding the carbons. At the start, this must allow the carbons to come together and then must pull them apart—called "striking the arc"—and it must also hold the arc at nearly a constant length. Fig. 330 shows the principle of a device for this. The carbon is at one end of a lever, and at the other end is the movable iron core of two solenoids A and B Excess current in the arc and hence in the "series" coil A pulls the core downward and separates the carbons. The "shunt" coil B acts oppositely.

The electric arc may be formed between any conductors which vaporise; at various times other materials than carbon, such as magnetite, have been used in commercial arc lamps. In the "flaming" arc, carbons impregnated with certain salts are used in order to increase the luminous efficiency.

461. Kinetic Theory of Heat Produced by an Electric Current.—For simplicity let us suppose that the electrons in a conductor are initially wholly at rest. When an electric force parallel to its length is applied to the conductor, each electron starts forward, but its forward motion is soon checked by a collision with an atom, and, since the rebounds of the various electrons will be in all directions, energy of undirected or random motion, that is, heat, will result. It is clear that the same effect must be continually taking place, whatever the actual state of motion at any moment may be. On the other hand, the random motion has in itself no tendency to return to directed motion. Thus heat is continually being produced at the expense of the current, and, in the steady state of the current, its energy is continually renewed by the agent (battery, dynamo, etc.), that keeps up the electric field. It can readily be shown that in each second a constant fraction of the energy of the current is changed into energy of random motion or heat. Now the kinetic energy of forward motion of the electrons is proportional to u^2 and, as we have seen, i is proportional to u. Hence the heat produced per second is proportional to i^2 and th is Joule's Law.

ELECTROLYSIS

462. Electric Conduction in Liquids and Electrolysis.—Some liquids act like metallic conductors, that is, the only change produced in the conductor by the passage of an electric current, be it either small or large, is due to the heat generated. Mercury and molten metals belong to the above class. But another class of liquids show, not only heat changes, but also chemical decomposition, when they are traversed by an electric current. Substances which are thus decomposed by an electric current are called *electrolytes*, and the phenomenon of chemical decomposition by an electric current is called *electrolysis*. Solutions of acids and salts, and molten salts are electrolytes. Fig. 331 represents a form of electrolytic cell which is convenient for showing electrolysis where gases are to be collected. A solution of hydrochloric acid ($HCl+H_2O$), is contained in the connected glass tubes, and the current enters and leaves the solution by the carbon terminals or electrodes A and K. The positive electrode is called the *anode*, and the negative the *kathode*. Upon the passage of an electric current, hydrogen escapes at the kathode,

and is collected in the glass tube above, and chlorine escapes at the anode. Since the chlorine gas is soluble in water, it does not appear in the tube until the water is saturated. The gases appear at the electrodes, and no decomposition appears in the body of the liquid. The part separated at the anode is called the *anion*, and the part at the kathode is called the *kation*. The term *ion* is used for either the anion or kation.

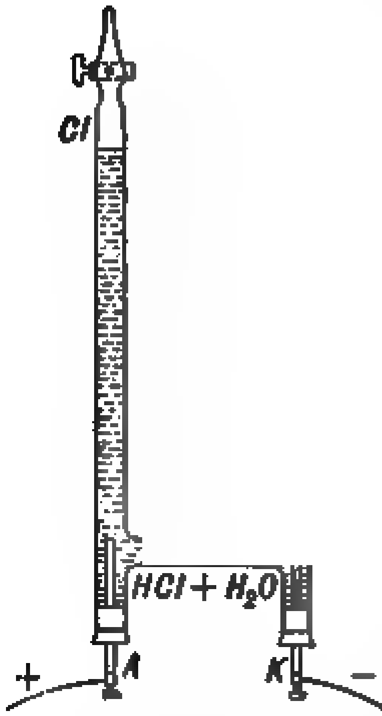


FIG. 331.

Conduction in electrolytes depends upon the formation of ions. It is assumed that when hydrochloric acid, HCl, is put in the water, it is ionized or dissociated, that is, it is broken into two parts or ions which have opposite electric charges, the hydrogen ion, *H*, which carries a positive charge and chlorine ion, Cl, which carries a negative charge. Under the action of the electric forces from the electrodes, these moving ions are directed into two opposite streams. It is the movement of these opposite streams of ions with their charges that constitutes the electric current. Electrolytic conduction is thus a convection process.

463. Secondary Changes in Electrolysis.—In the case described above, the ions appear at the electrodes and there is no intermediate chemical change. In many cases a secondary change takes place. In the electrolysis of a solution of sulphuric acid ($H_2SO_4+H_2O$), oxygen is obtained at the anode, and hydrogen at the kathode, there being two volumes of the hydrogen to one volume of the oxygen. The accepted explanation is that the sulphuric acid (H_2SO_4) is dissociated in the water into ions, the positive being H and the negative SO_4 or "sulphion." Under the directive action of the charged electrodes, a line of the hydrogen ions is drawn to the negative electrode, where they give up their charges and escape in bubbles. Similarly a line of sulphion ions is drawn to the positive electrode, and there they give up their charges. But the sulphion SO_4, cannot exist alone, and so it replaces the O in the water (H_2O), and O is released in

bubbles at the anode. The oxygen given off at the anode is thus the result of a secondary chemical action.

In the case of the passage of a current through a solution of copper sulphate $CuSO_4 + H_2O$, the electrodes being copper plates, the $CuSO_4$ is divided into the ions Cu and SO_4. The Cu is deposited on the kathode, and the SO_4, being released at the anode, combines with the copper of the anode, forming $CuSO_4$. We thus have the anode "wearing away," and the "electrolytic" copper deposited on the kathode.

464. Dissociation Theory.—The theory of electrolytic conduction outlined above assumes that every molecule of the substance before it goes into solution, is made up of two parts which are held together by their opposite electric charges, but when it is put in water, the binding force is decreased on account of the high dielectric constant of water (§§401 and 413), and so the substance is dissociated into ions. These ions are in constant motion in all directions until they are placed in the electric field between the two electrodes. By this electric field they are directed into two opposite streams, owing to their electric charges. When these streams of charged ions reach the electrodes, they lose their charges. This decomposition of the electrolyte in solution continues as long as the difference of potential between the electrodes is maintained by the external battery or dynamo.

To enumerate the many experiments and reasons for the above theory of electrolytic conduction is beyond the purpose of this presentation, but two significant facts may be mentioned. By themselves the constituents of an electrolytic solution are very poor conductors. Thus water freed from all its impurities is a "non-conductor"; and pure sulphuric acid is also a "non-conductor," but a solution of sulphuric acid in water is a good conductor, owing, as we have seen above, to dissociation or ionization. Again Kohlrausch has shown that the electrical conductivity of a dilute solution is directly proportional to the number of molecules of the salt or acid in the solution, and we are justified in assuming that all the molecules are ionized, and hence are carriers of electricity. The lowering of the freezing point of solutions, and the phenomena of osmosis, give added reasons for the dissociation theory of solutions, but the discussion of these last phenomena belongs particularly to physical chemistry.

465. Ohm's Law of Electrolytes and Polarization.—We have seen that for metals, the current is proportional to the electromotive force, that is, $i = CE$, where C is a constant for the circuit.

valencies, there will be two values for z. Thus iron for the ferrous salts has a valency of 2, and for the ferric salts a valency of 3, with the corresponding values for z as indicated in the table below.

Elements.	Atomic weight.	Valency.	Chemical equivalent.	Electro-chemical equivalent.
Chlorine..........	35.45	1	35.45	.0003672
Copper.............	63.6	2	31.8	.000329
Hydrogen...........	1.008	1	1.08	.00001044
Iron, ferric..........	55.9	2	27.95	.000289
Iron, ferrous........	55.9	3	18.49	.000193
Oxygen.............	16.0	2	8.0	.00008283
Silver...............	107.93	1	107.93	.001118
Zinc................	65.4	2	32.7	.000338

467. The Ionic Charge or "Atom of Electricity."—If we take the same number of grams of an element as the number denoting its atomic weight, and divide this by the valency, we get the gram-equivalent of the element. Thus the gram-equivalent of silver is (107.93)/1, and of copper (63.6)/2 or 31.8. It is evident from Faraday's laws that the quantity of electricity that deposits the gram-equivalent of one element will deposit the gram-equivalent of every other element. For silver this quantity is $107.94 \div .001118 = 96{,}550$ coulombs. Hence 96,550 coulombs will deposit the gram-equivalent of any element.

According to the dissociation theory, this charge is carried by the ions. Hence, if we can determine the number of ions in a gram, we can get directly from the above the electric charge carried by a single ion.

By methods given in special treatises on the kinetic theory of gases, the number of atoms in a gram of hydrogen has been calculated to be not far from 6×10^{23}. Hence the charge e per atom or ion for hydrogen is $e = 96{,}550/n$, or about 1.6×10^{-19} coulombs, or about 4.8×10^{-10} electrostatic units of electricity. We shall see later that this same charge e appears as the unit charge in the passage of electricity through gases (§ 564). This is the smallest quantity of electricity that we know, and all

other quantities appear to be multiples of this unit. Helmholtz as far back as 1881, noted the importance of this in the following remarkable words: "If we accept the hypothesis that the elementary substances are composed of atoms, we cannot avoid concluding that electricity is also divided into definite elementary portions which behave like atoms of electricity." These atoms of electricity we now call "*electrons*" (see §394).

468. The Voltameter or Coulombmeter.—The electrolytic cell gives us an accurate and convenient means of measuring electric currents for the calibration of instruments. An electrolytic cell arranged for measurement of currents is called a voltameter, or perhaps better a coulombmeter. The silver nitrate cell has been found to be capable of such accuracy in measuring currents that the international electrical congresses have adopted it as a convenient practical way of defining the "legal" ampere. Thus the ampere is defined for *practical* purposes as the current which flowing for one second through an electrolytic cell arranged according to directions fixed by law, deposits 0.0011180 grams of silver per second. Fig. 333 shows a silver nitrate coulombmeter.

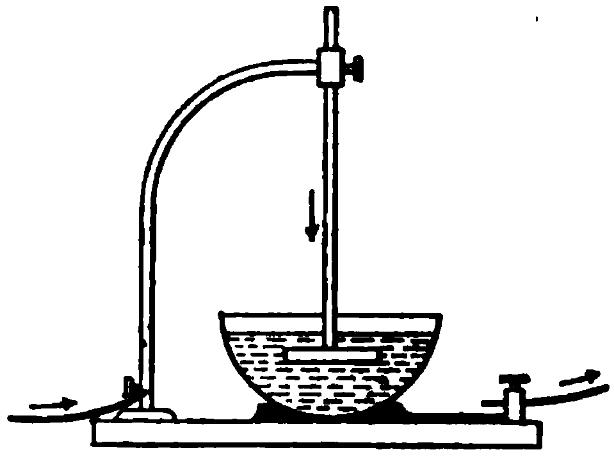


FIG. 333.

PRIMARY AND SECONDARY CELLS

469. Simple Voltaic Cell.—If a plate of copper and a plate of zinc are dipped into dilute sulphuric acid and are connected by a wire as shown in Fig. 334, an electric current is set up. The current flows through the wire from the copper to the zinc, and, in the solution from the zinc to the copper. The copper then forms the positive pole but the negative plate, and the zinc the negative pole but the positive plate of the cell. While the electric current flows, bubbles of hydrogen appear on the immersed part of the copper plate, and the zinc plate wears away, zinc sulphate being formed and going into solution. The above arrangement forms

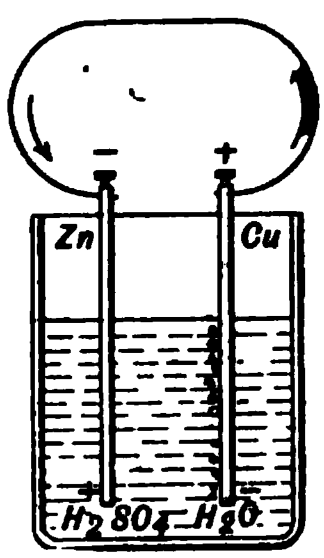


FIG. 334.

the voltaic cell discovered by Alexander Volta of Pavia, Italy, in 1800.

The first observations of what we now know as dynamic or sometimes as "galvanic" electricity, was made by Galvani, of Bologna in 1789. He discovered that when pieces of zinc and of copper were made part of a circuit with certain nerves and muscles of a freshly killed frog, that there was a contraction of the frog's muscles. Galvani recognised this as electrical, as he had produced similar effects with static electrical apparatus. The development of Galvani's discovery into the voltaic cell was made soon afterward by Volta.

The essential in Volta's cell is that there are two conductors and an electrolyte that acts chemically on one of the conductors more than on the other. The number of such voltaic combinations that are possible is indefinitely large. It is also possible to have a voltaic cell by substituting a suitable electrolyte for one of the metallic conductors. While many of these possible cells are interesting and important in the theory of the voltaic cell, only a few have practical value as generators of electric currents.

In Volta's simple cell, the current from the cell decreases very rapidly owing to the accumulation of hydrogen on the copper plate. The hydrogen causes a counter electromotive force of polarisation and also increases the internal resistance of the cell. To reduce or eliminate these polarization effects, and thus make a cell that will generate a more or less constant current, we have two general methods, the chemical, and the electrochemical method. In the chemical method, an oxidizing agent is placed round the negative plate, thus converting the hydrogen into water. An example of this is found in the Leclanché cell described later. In the electrochemical method there are two solutions, one around each plate, and the hydrogen combines with the solvent around the negative plate without freeing any polarising products. The Daniell cell, described later, gives a good example of this method. The above division of cells, evidently corresponds to a familiar division of cells into "single fluid" and "two fluid" cells.

470. Local Action.—Commercial zinc contains impurities, such as particles of iron and carbon, and when the zinc plate is immersed in dilute sulphuric acid, these impurities form with

the zinc of the plate, little local batteries. This "local action" consumes the zinc and covers the plate with a non-conducting film of gas. It has been found that by amalgamating the zinc with mercury, local action is largely eliminated.

471. Open and Closed Circuit Cells.—A cell which is used only at intervals and for short periods has time to recover from polarization either by diffusion of the gas or by the action of an oxidizing material. Such a cell is adapted for "open circuit" work, such as for ringing bells, etc., provided it has the additional property, that it does not deteriorate by "local action," or by harmful diffusions. The Leclanché cell is an example of a good open circuit cell. When more or less current is being used continuously, a "closed circuit" cell is needed. This should have no polarization. The Daniell cell, and the lead accumulator or storage cell (§474) are examples of good closed circuit cells.

472. Two Typical Voltaic Cells.—In this section we shall describe the Daniell and the Leclanché cells, since they are in very common use and also typical cells for closed and open circuit use.

One form of the *Daniell cell* is represented in Fig. 335. *Zn* is a rod of amalgamated zinc immersed in dilute sulphuric acid. This is in a porous cup *C*. Surrounding the cup is the glass jar *J* which contains a concentrated solution of copper sulphate and the copper plate *Cu*. The purpose of the porous cup is to keep the solutions from mixing and yet allow chemical action between the nascent hydrogen inside and the copper sulphate of the outside solution. When the copper and zinc poles are connected through an outside circuit *R*, an electric current flows through *R* from the copper to the zinc. On the inside the zinc unites with the sulphuric acid, forming zinc sulphate and freeing hydrogen. The hydrogen replaces the copper in the copper sulphate, and metallic copper is deposited on the copper plate. The reactions are represented as follows:

FIG. 335.

$$Zn + H_2SO_4 = ZnSO_4 + H_2, \text{ inside the porous cup;}$$

$$H_2 + CuSO_4 = H_2SO_4 + Cu \text{ on the outside of the porous cup.}$$

The *gravity cell*, represented in Fig. 336, differs only from the Daniell cell in that the separation of the two solutions is maintained by their different densities. The dense copper sulphate occupies the bottom of the jar, and the lighter acidulated solution rests above it. Under suitable conditions the solutions do not mix to an extent that affects seriously the action of the cell.

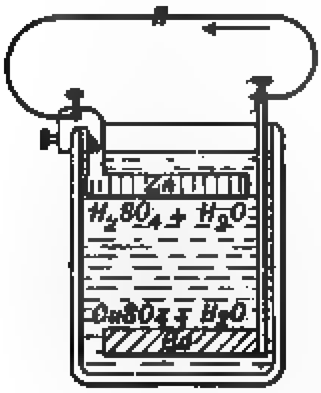


Fig. 336.

The e.m.f. of a Daniell or a gravity cell is ordinarily about 1.08 volts. Since there is no polarisation, the cell gives a constant current. The internal resistance of the cell is, however, comparatively large, so that from a cell of ordinary size only about an ampere can be taken. The cell has been largely used in telegraphy where constant currents are needed.

The *Leclanché cell* is a single fluid cell, using a solution of sal-ammoniac, and the plates are zinc and carbon. The carbon is enclosed in a porous cup and packed around with manganese dioxide and broken carbon. The sal-ammoniac solution diffuses through the porous cup to the carbon. The action is described as follows: The zinc unites with the sal-ammoniac (NH_4Cl), forming $ZnCl_4$ and NH_3 and H. The hydrogen unites with the MnO_2 forming M_2O_3 and H_2O. The initial e.m.f. of this cell is about 1.5 volts. This falls off more or less when the current flows, as the hydrogen is not oxidized by the manganese dioxide as rapidly as formed. The cell recovers, however, when left on open circuit.

Fig. 337.

The *"dry cell"* which is used so extensively for spark coils, electric bells, etc., may be regarded as a form of the Leclanché cell. The zinc is in the shape of a cylindrical cup which forms the vessel for the cell. The carbon rod and the oxidizing dioxide of manganese are at the center of this cup, and are surrounded with a packing of some absorbing substance such as saw dust. This

Fig. 338.

is saturated with sal-ammoniac solution, and the cup is sealed with pitch to prevent evaporation.

473. Standard Cells for E.M.F. Determinations.—In calibrations with the potentiometer (§455) it is necessary to have a "normal" or "standard" cell of known and constant e.m.f. The two cells used universally for this purpose are the cells devised by Latimer Clark and by Edward Weston. A form of the Clark cell is shown in Fig. 339. The positive pole is mercury (Hg), in contact with a paste of mercurous sulphate (Hg_2SO_4), and the negative pole is zinc in contact with the solution which is zinc sulphate. When this cell is made strictly according to the specifications fixed by the national physical laboratories, it has an e.m.f. of 1.434 volts at 15°C. and, for a temperature t, an e.m.f. of [$1.434-0.0012\,(t-15)$] volts.

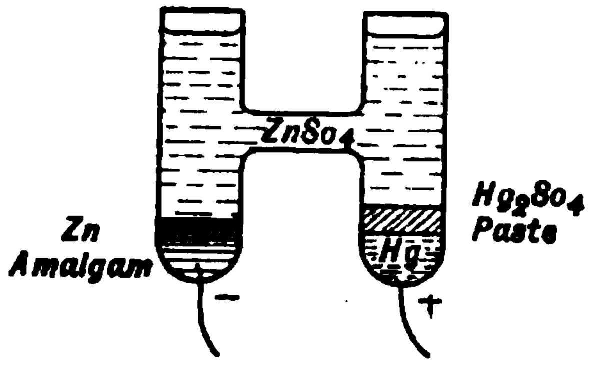


FIG. 339.

The Weston cell is exactly like the Clark cell except that the zinc is replaced by cadmium, and the zinc sulphate by cadmium sulphate. Its e.m.f. in the standard form is 1.0190 volts, and it has the great advantage of having practically no change of e.m.f., with temperatures. No appreciable current should be taken from a standard cell, as the accompanying chemical actions cause more or less permanent changes in the cell and its e.m.f.

474. Storage Cells.—It was noted that the e.m.f. of polarization in an electrolytic cell is due to the gases or other chemical products formed on the electrodes. Thus in an electrolytic cell with platinum electrodes in dilute sulphuric acid, oxygen collects on the anode and hydrogen on the cathode, that is, we get the equivalent of a voltaic cell, with plates of oxygen and hydrogen. When the external current of this electrolytic cell is broken and the cell is joined to a circuit containing a galvanometer, we get a current from the oxygen "pole" to the hydrogen "pole," that is, opposite to the current which produced the electrolysis. A cell thus formed by electrolytic action is called a *secondary* cell, in distinction from voltaic cells, which are called "primary" cells. Secondary cells are perhaps more commonly called *storage cells* or electric *accumulators*. It is however to be noted that the energy "stored" or "accumulated" in a storage cell is chemical energy and not electrical energy.

The current from the "gas" storage cell described above is of short duration, as the gas layers are rapidly diffused. In 1860 Planté discovered that, by using lead plates in a sulphuric acid solution, a secondary battery could be formed of large capacity and one which could be discharged days or weeks after the time of charge. In the Planté cell, the surface of the anode becomes coated with red oxide of lead, PbO_2, and the hydrogen escapes at the cathode. If, after being "charged," the plates are connected through an external circuit, the chemical action of charging is reversed and a current is taken off in the reverse direction to the charging current. This continues until the products of the electrochemical decomposition are consumed. Planté found that the capacity of the lead plates could be greatly increased by a system of charging, discharging, and reversing the charges, this "forming" process often taking many hours. Faure found that he could shorten the time of forming a cell by covering the anode with a paste of peroxide of lead. The lead storage cell has a normal e.m.f. of about 2 volts, and this e.m.f. remains almost constant until the cell nears the discharged condition. The internal resistance is low, and the current output is large.

For an account of different types of storage cells, and of the chemical and electrical transformations involved in their action, special treatises must be consulted.

475. Theory of the Voltaic Cell.—The decomposition of the electrolyte in the voltaic cell is the same as that in an electrolytic cell; but, in the case of the electrolytic cell, the e.m.f. between the electrodes and the energy of the process is maintained by an outside source, while in the voltaic cell the e.m.f. is produced in the cell itself. The origin of this e.m.f. in the voltaic cell has been one of the debated problems of physics for over a century. There have been two theories, the contact theory due in its original form to Volta, and the chemical theory which was held early by Faraday and others. Both theories have naturally been modified in various ways as new facts have been discovered. According to Volta, there is a difference of potential between two unlike bodies, due merely to their contact. By means of a sensitive electroscope Volta showed that, when plates of zinc and copper are brought into contact in air and separated, the

zinc becomes positively electrified and the copper negatively electrified. As the result of his experiments, Volta made a series in which the metals are arranged in order so that each metal is positively electrified when placed in contact with a metal lower in the series. Volta's series was zinc, lead, tin, iron, copper, silver, gold, carbon. Later observers, using modern sensitive electrometers, have confirmed the essential facts of Volta's fundamental experiment. If two metals given in this series form the plates of a voltaic cell, the first in the series forms the positive plate and the second the negative plate.

The interpretation of Volta's contact experiment has been the point of controversy. On the chemical theory, the potential of "contact" is due to oxidation by films on the metals, the electrical transfer being due to this chemical action. Since all metals have been in air, and such invisible films are persistent, it has been impossible to arrange test experiments free from all objection. But it is found that metals boiled in a mineral oil at a high temperature, which presumably removes any such films, show no difference of potential on contact, and this is urged in favor of a chemical origin of the contact e.m.f. It should be noted that even if we hold that the source of the e.m.f. is to be sought in the contact of unlike bodies, the keeping up of the electrical transfer, that is, of the current, is due to the chemical work. Hence, whether we think of the chemical action as occasioned by the contact e.m.f., or think of the e.m.f. as due to chemical action, the study of the energy transformations is one of chemical energy. The phenomena of these transformations in the voltaic cell are of a very complicated nature. For further discussion of this very interesting subject the student is referred to Nernst's *Theoretical Chemistry*.

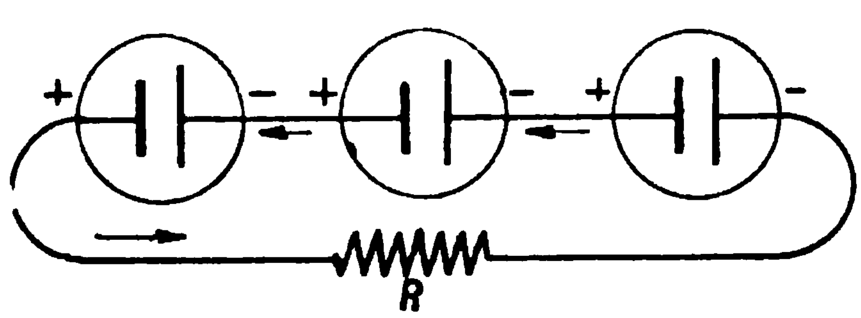


FIG. 340.

476. Series and Parallel Arrangement of Cells.—Cells are "in series" when the positive pole of each cell is joined to the negative pole of the next cell. The circuit is completed by a conductor of

resistance R joined between the positive and negative poles at the ends of the series (Fig. 340). The total e.m.f. is the sum of the e.m.f.'s of the n cells, or is nE. The internal resistance is nr, so that the total resistance of the circuit is $R+nr$. The total current is then, by Ohm's law,

$$I=\frac{nE}{R+nr}$$

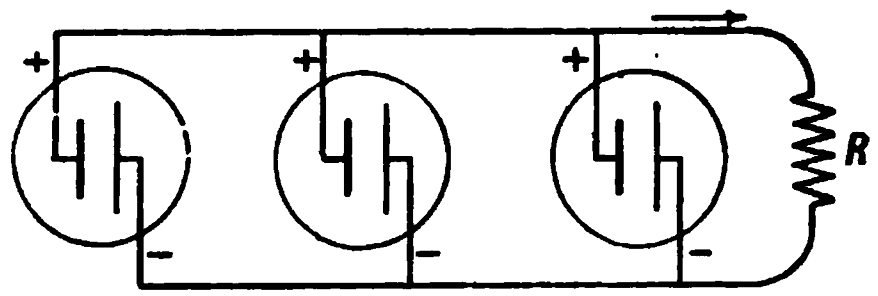


FIG. 341.

Cells are "in parallel," when all the positive poles are joined together and all the negative poles are joined together, (Fig. 341). The cells are thus equivalent to a single large cell with an internal resistance of r/n, and the e.m.f. is E, that of a single cell. The current through an external resistance R is then

$$I=\frac{E}{R+r/n}$$

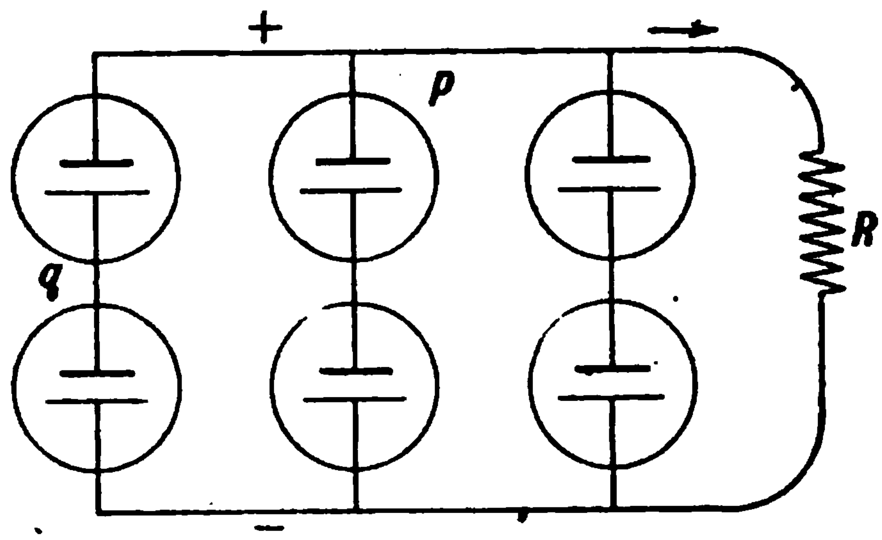


FIG. 342.

Cells can also be joined in a series-parallel arrangement of p rows in parallel, each row having q cells in series (Fig. 342). The total number of cells is $n=pq$. For each row the e.m.f. is qE, and the internal resistance is qr. For the p rows in parallel, the in-

ternal resistance is qr/p, and the e.m.f. is qE. Hence the current through R is

$$I=\frac{qE}{R+qr/p}=\frac{pqE}{Rp+qr}=\frac{nE}{Rp+qr}$$

The maximum current is obtained when the cells are arranged so that the internal resistance is made as nearly equal to the external resis'ance as possible.

This is shown as follows: From the equation $I=(nE)/(Rp+qr)$, it is evident that I is a maximum when $Rp+qr$ is a minimum. Write this in the form $(\sqrt{Rp}-\sqrt{qr})^2+2\sqrt{Rpqr}$. The term $2\sqrt{Rpqr}$ is a constant for a given external resistance and a given number of cells. Hence the value of the expression is least when $(\sqrt{Rp}-\sqrt{qr})^2=0$, that is, when $Rp=qr$, or $R=qr/p$. But qr/p is the battery or internal resistance. Hence the current is a maximum when the internal resistance is equal to the external resistance. Half of the energy of the current in this last case goes into heating the cells.

THERMOELECTRICITY

477. Thermoelectric Currents.—If a copper and an iron wire are joined to form a circuit (Fig. 343) and one junction of the two metals is heated, an electric current is set up. The current flows from iron to copper across the cold junction for ordinary ranges of temperature. Such an arrangement of metals forms a *thermocouple.* This phenomenon of thermoelectricity was discovered in 1821 by Seebeck who showed that such currents were produced by the unequal heating of junctions in circuits of any dissimilar metals. The electromotive forces produced in this way are very small, only a small fraction of a volt per couple in the most favorable combinations. (See §479.)

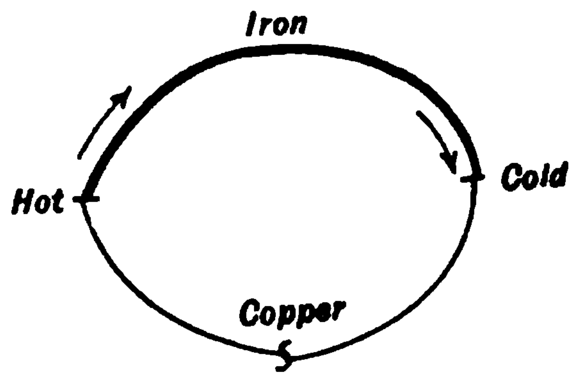


Fig. 343.

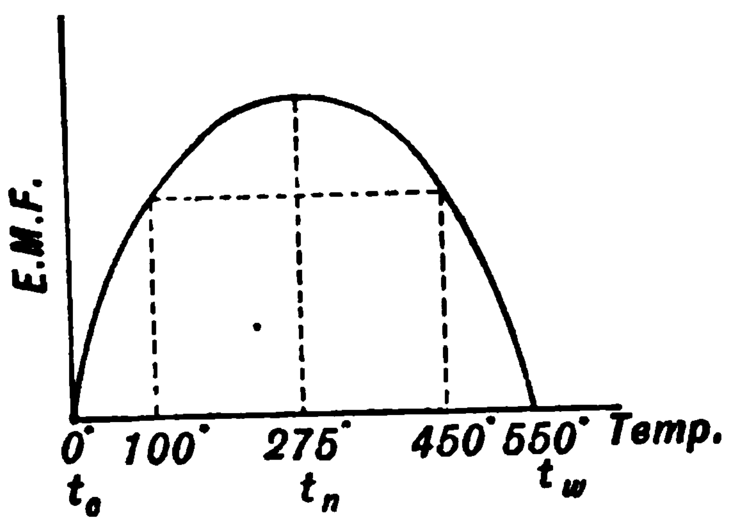


Fig. 344.

If one junction of iron-copper couple is kept at 0° C., and the temperature of the other junction is raised, the e.m.f. increases until a temperature of about 275° C. is reached. This is called the *neutral temperature* for the couple. The e.m.f. now decreases and becomes

sero when the temperature of the hot junction is about 550° C. Beyond 550° the e.m.f. is in the inverse direction, that is, from copper to iron across the cold junction. If the temperature of the cold junction is raised, say to 100°C., the temperature of inversion is lowered to 450°C., that is, the mean of the temperatures of the cold and hot junctions at inversion is equal to the neutral temperature. The general form of the curve for the temperature and the e.m.f. is shown in Fig. 344. In this t_c and t_w are the temperatures of the cold and warm junctions and t_n the neutral temperature. It has been found for most cases to be a parabola, and so the curve can be determined for a particular couple, if the e.m.f. is known for three suitable temperatures.

From the above it is seen that the thermal e.m.f. depends upon:

(*a*) the metals of the couple;

(*b*) the difference of temperature of the junctions;

(*c*) the mean temperature of the junctions.

478. Effect of Intermediate Metals.—In the circuit *ABCD A* (Fig. 345) consisting of the two metals *B* and *D* with junctions *A* and *C*, at temperatures t_1 and t_2 respectively, let the junction *C* be broken, and let a third metal *X* be introduced. If the new

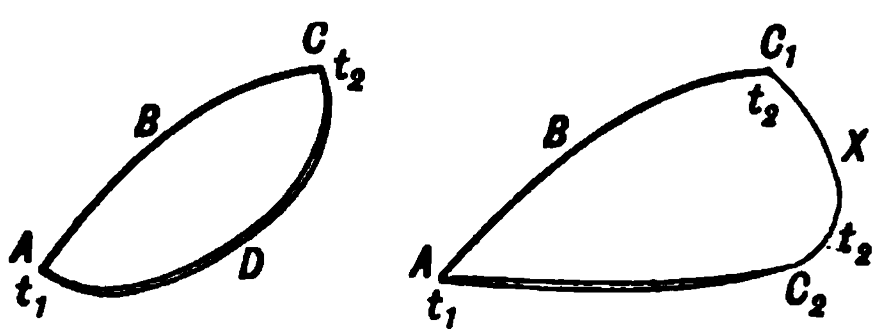


FIG. 345.

junctions remain at the temperature t_2, experiment shows that the e.m.f. of the circuit is not changed.

From this we see that the junction of two metals for a thermo-couple can be made either directly or by solder; further that we can connect a galvanometer in the circuit of a thermo-couple by intermediate wires without affecting the e.m.f., provided the temperature of the junctions in the connecting circuit are kept constant. If the temperatures of the new junctions are not uniform, the effect is that of introducing additional thermo-couples into the circuit.

479. Thermo-electric Power and the Thermo-electric Diagram.—Thermocouples have found a large use for measuring differences of temperature. For this use we want to know *the e.m.f. per degree temperature difference.* This is called *the thermo-electric power* of the couple, and, as we have seen, it depends upon the mean temperature of the junctions. In Fig. 346, the

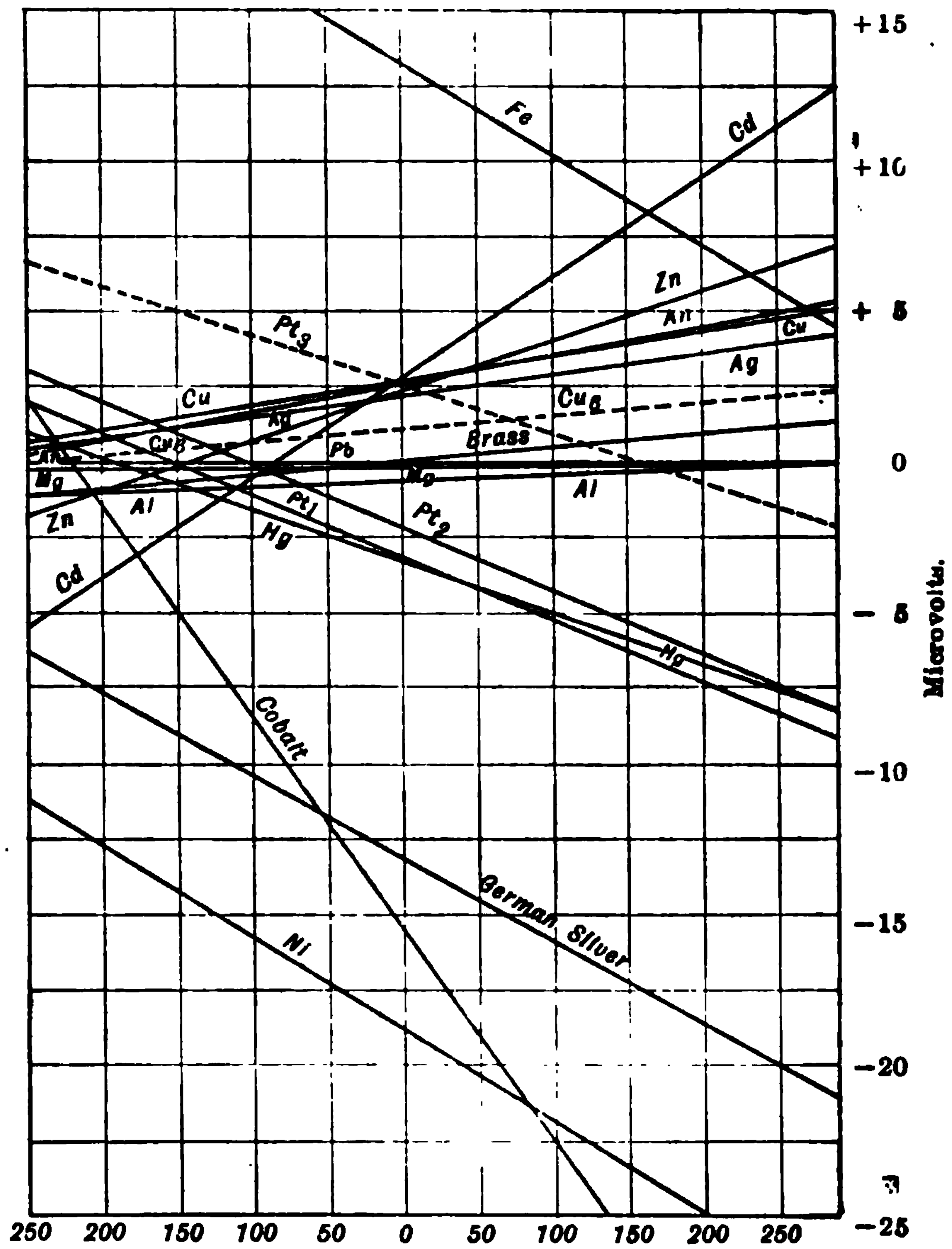


FIG. 346.

ordinates represent the thermo-electric powers of a number of metals referred to lead and the abscissas represent the *mean* temperatures. These experimental curves of the thermo-electric powers are practically straight lines within these limits of temperature. To get the thermo-electric power of any couple for a given mean temperature, for example, an iron-copper couple for a mean temperature of 50°, we read the length of the ordinate between the iron and copper lines at the abscissa distance of 50°. In this

case, it is about 8.7 microvolts per degree difference in temperature. From the points where the lines of the two metals intersect, we get the neutral temperature for the couple.

480. Peltier Effect.—Peltier discovered in 1834 that if a current is sent through the circuit of a thermo-couple, heat is given out at one junction and absorbed at the other junction. If the current is reversed, the junction that was heated is now cooled and the other is heated. This effect is due to the fact that at one junction the current opposes the potential difference between the two metals, and hence work is done there by the current, and this electric energy appears as heat. At the other junction, the potential difference of the two metals acts with the current and there is cooling. To show this heating and cooling effect Peltier made a cross of bars of antimony and bismuth and joined a battery and a galvanometer as shown in Fig. 347. When the current flows from the antimony, *DC*, to the bismuth, *CB*, there is a cooling, as shown by the thermo-couple circuit *CAGE;* when the current is reversed, that is, so as to flow from bismuth to antimony, the galvanometer shows a heating of the junction. Tyndall demonstrated the same phenomenon by passing a current through an ordinary thermo-pile and upon breaking the current, quickly introduced a galvanometer in circuit. The galvanometer showed an inverse current in the thermo-couple, corresponding to the unequal temperatures at the two sets of junctions from the Peltier effect of the first current.

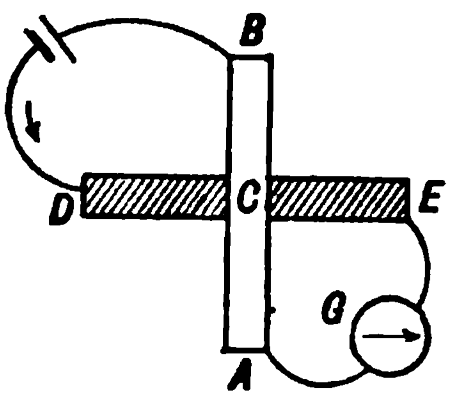


Fig. 347.

481. Thomson Effect.—Lord Kelvin has shown that if an electric current is passed through a bar along which there is a flow of heat, there is an absorption or generation of heat in the bar, which depends upon the direction of the current and the nature of the metal. Thus if an electric current passes along a copper bar from the cold to the warm part, the copper is cooled. If the current is reversed so as to pass from the warmer part to the colder, the copper is warmed. In iron the Thomson effect is opposite to that in copper. The effect in lead is practically zero, and hence lead is commonly used as the comparison metal in thermo-electric diagrams.

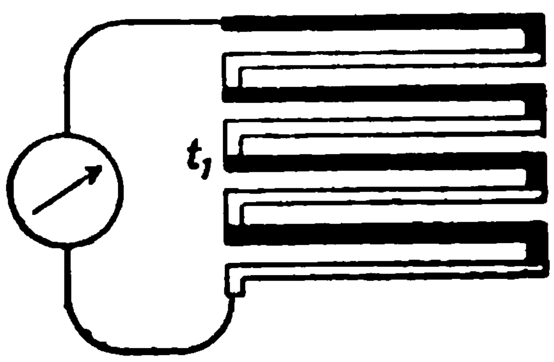


Fig. 348.

482. Applications of Thermo-couples.—As generators of electric currents, thermo-couples have little use owing to their small e.m.f., and their comparatively high internal resistance. But they have found a large and valuable use for temperature determinations, particularly for very small differences of temperature, for very high and very low temperatures, and for the temperatures of bodies inaccessible to ordinary thermometers.

For small temperature differences, thermo-couples are often arranged in the form of a *thermo-pile.* This consists of alternate bars of the two metals arranged in a zigzag order (Fig. 348) and built into a cubical block or

pile. The even-numbered junctions form one face of the pile, and the odd-numbered junctions the opposite face. A difference of temperature between the two faces thus produces a series of e.m.f.'s, the addition of which forms the total e.m.f.

One of the most sensitive thermo-electric arrangements for detecting small differences of temperature is Boys' radio-micrometer. This consists of a single bismuth antimony couple, the circuit of which is completed by a loop of copper wire, which is suspended by a quartz fiber between the poles of a strong magnet (Fig. 349). The loop of wire thus forms the coil of a d'Arsonval galvanometer, and gives a very sensitive means of detecting a thermal current from the antimony-bismuth couple. The couple is diamagnetic and so has to be screened magnetically by a soft iron block.

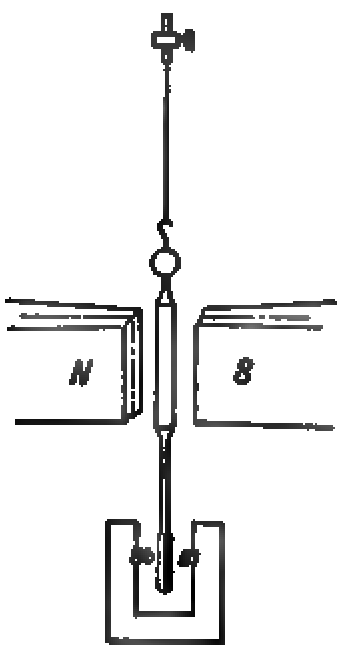


FIG. 349.

FIG. 350.

With this instrument it is said that the heat of a candle five hundred yards away can be detected, or a rise in temperature of less than one millionth of a degree.

Duddell has made an important application of the Boys' radio-micrometer in a thermo-galvanometer for measuring small oscillatory currents, such as are used in telephony and in the antennæ of wireless telegraphy. The current heats a coil R (Fig. 350) which is placed just below the suspended thermo-couple AB. The heat generated in R is determined by the deflections of the coil of the radio-micrometer, and thus a measure of the current is obtained. Since the heating does not depend on the direction or the frequency of the current, this galvanometer can be usedfor oscillatory currents of any period. Instruments of this kind have been made sensitive to 2.2×10^{-7} amperes.

For measuring temperatures, from 600° to 1600°C., a thermo-couple of platinum and an alloy of platinum with ten per cent. rhodium is found most

satisfactory. The LeChatelier "thermo-electric pyrometer" consists of such a thermo-couple with a suitable millivoltmeter graduated with a temperature scale. For lower temperatures, copper-constantan couples are much used.

ELECTROMAGNETS AND MAGNETIC INDUCTION

483. Electromagnets.—A coil of insulated wire around an iron core forms an *electromagnet*, when an electric current flows in the coil. The direction of the magnetization is determined by the direction of the magnetic field of the coil. Thus in a helix AB (Fig. 351*a*) the current flows anti-clockwise when we look at the face B, and hence that end is a magnetic N face with lines emerging as indicated (§427). The same helix with an iron core (Fig. 351*b*), shows the same direction of magnetization, since the molecular magnets of the iron tend to line up in the direction of the magnetic field of the coil. Electromagnets are used instead of permanent magnets, (*a*) where very strong fields or very strong poles are needed; and (*b*) where it is desired to vary the strength of the magnet or to reverse its polarity. The latter can be done by varying or reversing the magnetizing current. The common uses of electromagnets are to produce magnetic fields as in dynamo machines, and to exert forces as in magnets for lifting loads, and in signaling apparatus (telegraphy, electric bells, etc.).

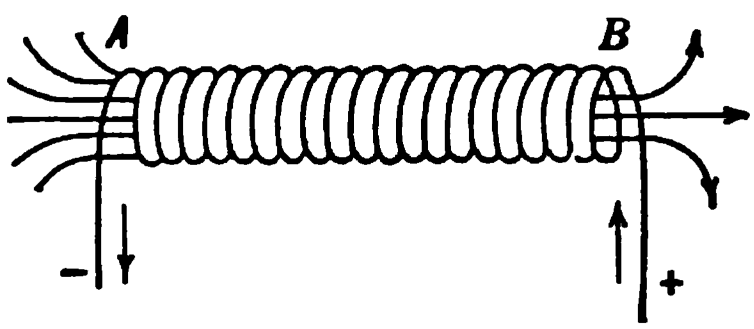


FIG. 351*a*.

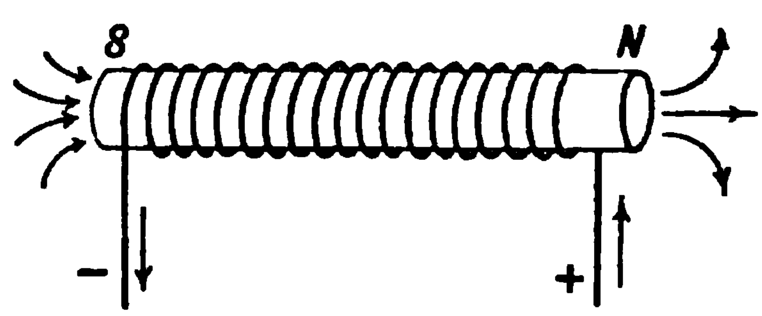


FIG. 351*b*.

The design of an electromagnet involves a study of the magnetic properties of the iron core, and also a calculation of the magnetic field of the coil. The magnetic field of the coil depends upon the dimensions of the coil and the current, as has already been indicated for some special cases (§430).

484. Magnetization of Iron.—When a piece of iron is placed in a magnetic field, it becomes magnetized, but only for special

forms of the specimen is the direction of the magnetization the same as that of the external magnetic field. The induced poles act against the external field and so have a demagnetizing action. In some cases, as in that of a short bar, this demagnetizing action is very strong. In general the intensity and direction of the resultant field, and hence the magnetization, does not admit of calculation. In two cases it is possible to calculate the resultant magnetizing force acting on the iron—(1) that of an ellipsoid of revolution with an axis in the direction of the field and (2) that of an anchor ring or toroid.

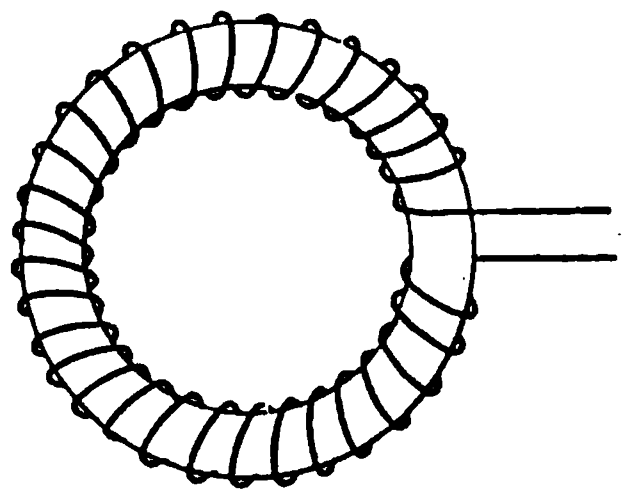


FIG. 352.

(1) The case of the ellipsoid is attained closely enough for practical purposes by using a cylindrical rod with its length 400 or 500 times its diameter, the axis of the rod being in the direction of the field. The uniform field is produced by a solenoid, which is somewhat longer than the rod. The poles induced in the rod are so distant that they do not appreciably change the direction of the field near the middle of the solenoid, and there the direction of the field and of the magnetization coincide.

(2) In the second case, that of the anchor ring, the magnetic field is produced by an endless solenoid wound on the ring as core (Fig. 352). There are no free poles in the ring and hence the field of the solenoid and the magnetization have the same directions.

There are two ways of expressing definitely the magnetic condition of iron or other magnetized substance. In the first way, we use a quantity called "*the intensity of magnetization,*" this being represented by the letter "I." In the second way, we use a quantity called "the magnetic induction," this being represented by the letter "B." We shall define "I" and "B" in the following sections, and describe the methods of testing the magnetic qualities of iron.

485. Intensity of Magnetization. IH Curve of Magnetization.— Consider a small right cylinder of the iron of volume v magnetized parallel to its axis. Let its length be l and the area of its end s. If M is its magnetic moment, then M/v, or the *magnetic moment*

per unit volume, is called *the intensity of magnetization*, I, that is, $I=M/v$.

If m is the strength of the pole, $M=ml$ and $v=sl$. Hence $I=M/v=ml/sl=m/s$, where m is the pole strength for the sectional area s. Hence I, the intensity of magnetization, is equal to *the pole strength per unit area* of a cross-section at right angles to the magnetization.

In the case of a long thin cylinder, placed parallel to the field, the magnetization is in the direction of the field, and hence the free polarity is wholly on the ends. In this case, the intensity of magnetization is $I=m/\pi r^2$, where m is the pole strength and πr^2 is the area of the cross-section.

We can measure the pole strength m by the deflection of a magnetic needle as described in the next section. The area πr^2 is known, and thus I can be calculated. The corresponding value of H, the magnetizing force, is found from the current and the constants of the solenoid. The relation between I and H can now be shown by drawing a curve with the values of H as abscissas, and of I as ordinates. Fig. 353 shows such an IH *curve of magnetization* for Norway iron. At first, from O to A, the curve rises very slowly, and then it rises rapidly to a saturation bend B. The part AB is almost a straight line. From B it rises very slowly for large increases of the field intensity H.

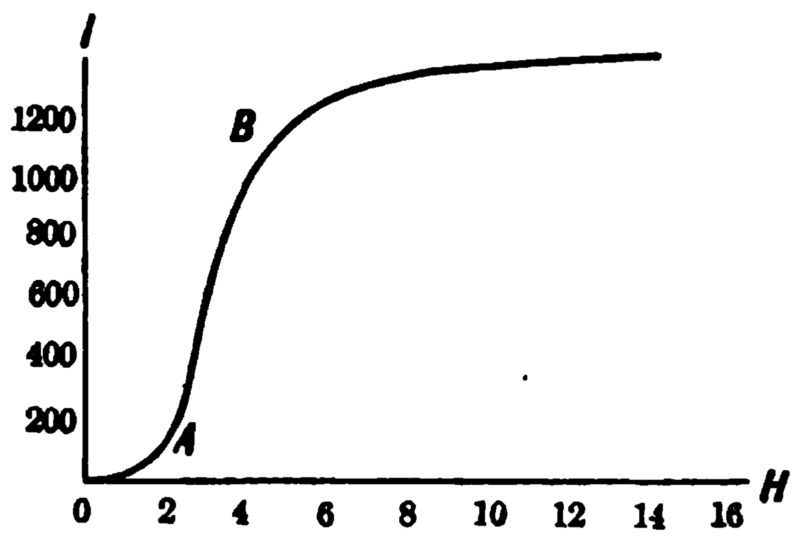


FIG. 353.

The explanation of this curve is simple. During the first part, that is from O to A, the groups of little magnets, formed by their mutual attractions, are being broken up. As soon as these groups are broken up, the elementary magnets fall rapidly into the line of the field, so that at B, almost all of them are pointing in the direction of the field.

The ratio of the intensity of magnetization to the intensity of the magnetizing field is called *the magnetic susceptibility*, k, of the material. Hence

$$k=I/H$$

If the magnetic susceptibility k were a constant, the IH curve of magnetization would be a straight line. It is evident from the experimental curve that k depends not only on the magnetic substance, but also upon its intensity of magnetization.

In Fig. 354, we have IH curves for several materials, showing how these substances differ in their magnetic susceptibility. The results are only approximate since the magnetic properties of a material change with treatment.

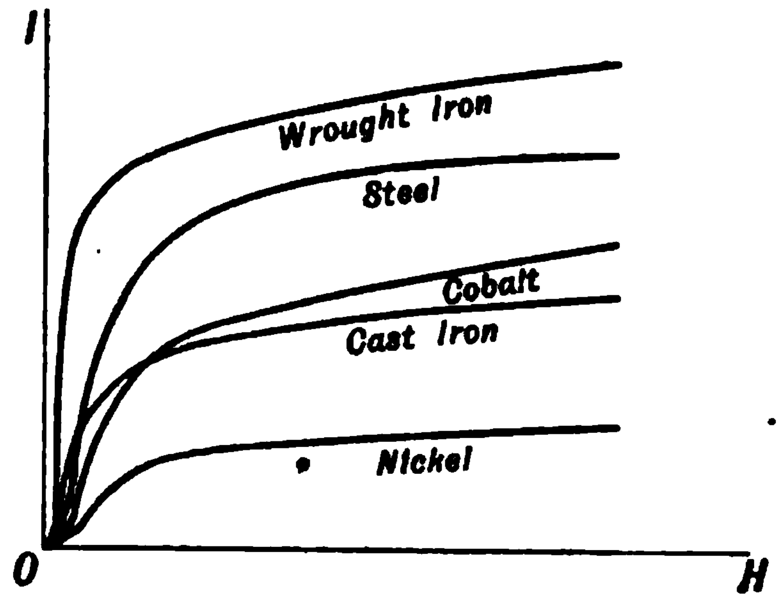


FIG. 354.

486. Magnetometric Method for Obtaining Magnetization Curves.—The method as used by Ewing is shown in Fig. 355a. The iron to be tested is a long thin wire, ns, in a magnetizing solenoid, AB. This is placed vertically with the upper end of the wire ns on the same level as a small magnetometer, M, and either east or west (magnetically) from M. The magneto-

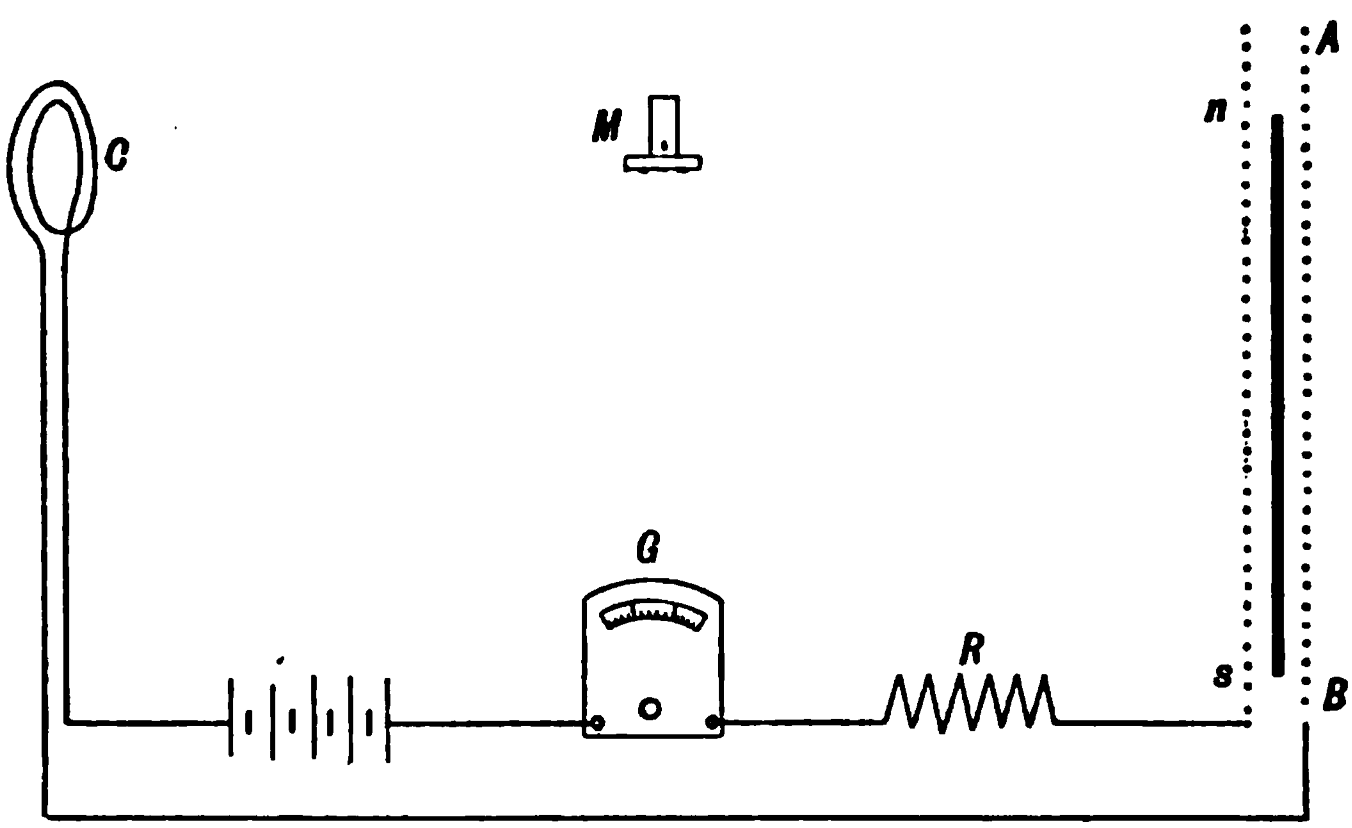


FIG. 355a.

meter consists of a short magnetic needle suspended by a fine quartz or silk fiber, and supplied with a mirror for reading scale deflections with a lamp or telescope, (Fig. 355b). The horizontal field R acting on M, due to the magnetized wire ns, is calculated as follows: The intensity at M due to $+m$ is m/d^2_1; the intensity at M due to $-m$, resolved horizontally is $-m/d^2_2\ (d_1/d_2)$. Hence the total horizontal intensity at M due to ns is,

$$R=\frac{m}{d^2_1}-\frac{m}{d^2_2}\left(\frac{d_1}{d_2}\right)=\frac{m}{d^2_1}\left\{1-\left(\frac{d_1}{d_2}\right)^3\right\}$$

Substituting for m its value $\pi r^2 I$, we get

$$R = \frac{\pi r^2 I}{d^2_1}\left\{1-\left(\frac{d_1}{d_2}\right)^3\right\}$$

If H is the intensity of the field directing the magnetometer "northward," then by the tangent law (§382), we have

$$\frac{R}{H} = \tan\phi = \pi r^2 \frac{I}{H}\left\{1-\left(\frac{d_1}{d_2}\right)^3\right\}\left(\frac{1}{d^2_1}\right)$$

or

$$I = \frac{(d^2_1)\, H \tan\phi}{\{1-(d_1/d_2)^3\}\pi r^2}$$

We can thus calculate I from the deflections of M.

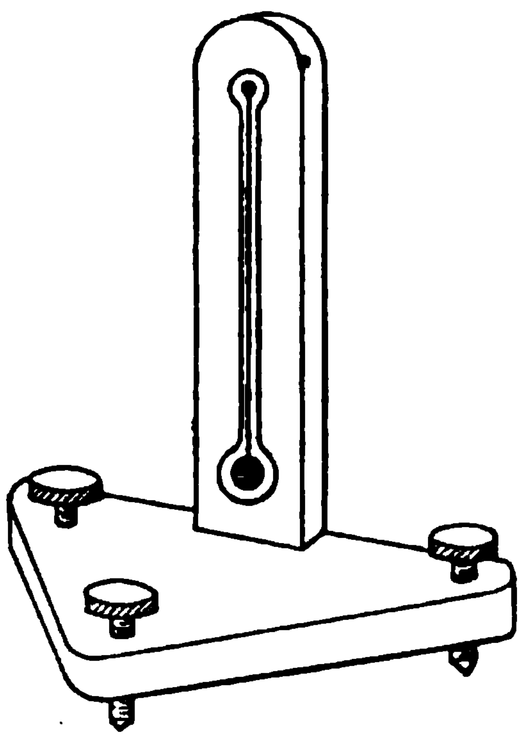


FIG. 355b.

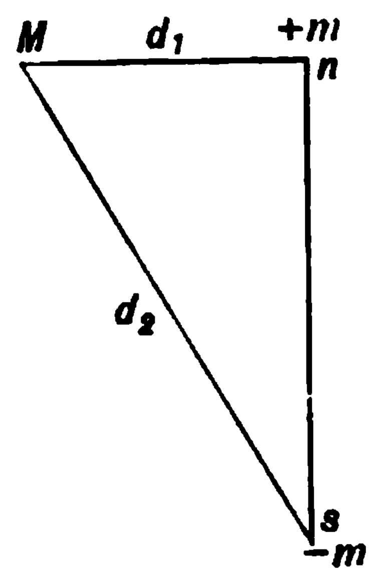


FIG. 355a.

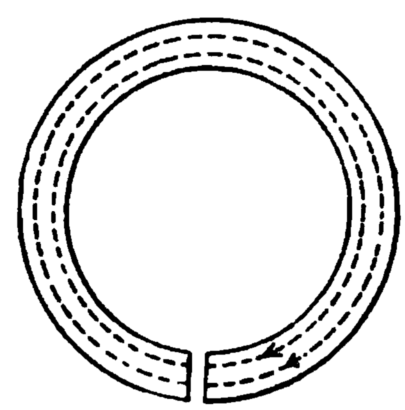
To get the magnetic intensity due to ns, independently of the action of the solenoid AB on M, a compensating coil C carrying the same current as AB is placed so that M is not deflected by the currents in AB and C when the iron ns is not in AB. The field H, is calculated from the formula for the solenoid, $H = \frac{4\pi n i}{l}$ (§430). The current i is measured by the galvanometer or ammeter G, and is increased or decreased by means of the rheostat R. From the values of H and corresponding values of I a curve is plotted.

487. Magnetic Lines closed Loops. Lines of Magnetization. Magnetic Induction.—We have seen that the space about a magnet is a region of magnetic stresses which are shown by the lines of magnetic force. As we have described these lines, they emerge at the north pole and enter at the south pole, their whole course, so far as it has been described up to this time, being in the air or other non-magnetic medium. In this respect the mag-

netic lines seem, at first sight, to be exactly similar to electrostatic lines of force, which, as we have seen, start from positive electricity and terminate in an equal quantity of negative electricity, and do not continue into the conductors (§398). That is, electrostatic lines are not closed loops. We shall show now that the *magnetic lines* of a magnet are continued into and through the iron and *form closed loops.*

It has been already seen that the magnetic lines around currents are closed loops. Thus around a linear circuit we have circular lines, and about a solenoid lines which emerge from the *N* face, enter the *S* face and complete the loop through the solenoid (see Figs. 300 and 301, §427).

Consider now the case of an anchor ring magnetized by a helix or solenoid which is wound on the ring as core. Here there are no poles, and so there is no magnetic field outside of the ring. But the ring is magnetized, as could be seen by cutting the ring into sections. Each section, as in the case of the broken magnet (§366), would be found to be a magnet. Suppose a narrow gap to be cut normally across the ring (Fig. 356). On one side of this gap we find a *N* pole and on the opposite side a *S* pole. If I is the intensity of magnetization, and s the area of the section, the pole strengths at the gap are $+Is$, and $-Is$. Hence, $4\pi Is$ lines of force (§374) cross the gap from the positive to the negative side. Since this will hold for any and every gap however narrow, we are led to think of these $4\pi Is$ lines as continuous lines extending round the ring. In an air gap these are lines of force, but in the metal, they are called *lines of magnetization*, and we may suppose them to be due to the lined-up elementary magnets of the metal. But it is to be noted that a line of force in the gap is the continuation of the line of magnetization in the metal.



$-Is+Is$

FIG. 356.

In addition to these lines of magnetization there are lines of force, H per square centimeter, due to the magnetizing solenoid. In the present case the magnetizing field and the magnetization coincide in direction, that is, the lines H and $4\pi I$ (per square centimeter) are to be added algebraically. The total number of lines per square centimeter is therefore $H+4\pi I$. This sum is

called *the magnetic induction*, and is represented by the letter B. That is,

$$B=H+4\pi I$$

488. Magnetic Flux.—The total number, N, of magnetic lines that pass through an area, S, normal to the field is evidently equal to the product of the area and the magnetic induction, that is,

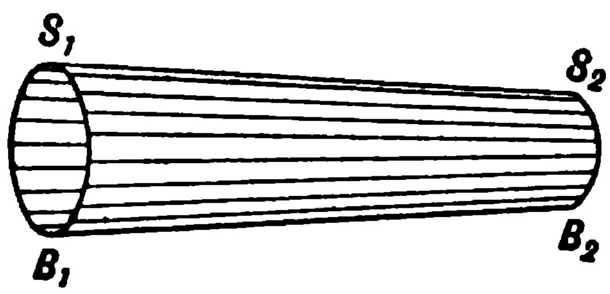


FIG. 357.

$$N=BS$$

A "tube" bounded by magnetic lines (Fig. 357) evidently has the property that the product of the area of any *normal* section by the magnetic induction, BS, is a constant. This is exactly analogous to the steady flow of a fluid in a pipe, where the product of the flow per unit section and the area of the section is a constant. Hence, Maxwell and others have used the term *magnetic flux* or flow of magnetic induction for the product $BS=N$. The idea suggested by the term is very helpful in discussing the magnetic induction in transformers and other electromagnetic apparatus.

489. BH Curves of Magnetization.—From the IH curve of magnetization of Fig. 353 and the relation $B=H+4\pi I$, we can calculate B and draw a curve showing the relation between B and H (Fig. 358). The BH curve is in many ways more valuable than the IH curve, particularly in connection with apparatus for electromagnetic induction. In such apparatus we need to know $BS=N$, the total magnetic lines or flux, and also the magnetizing field H, which will produce this magnetic flux.

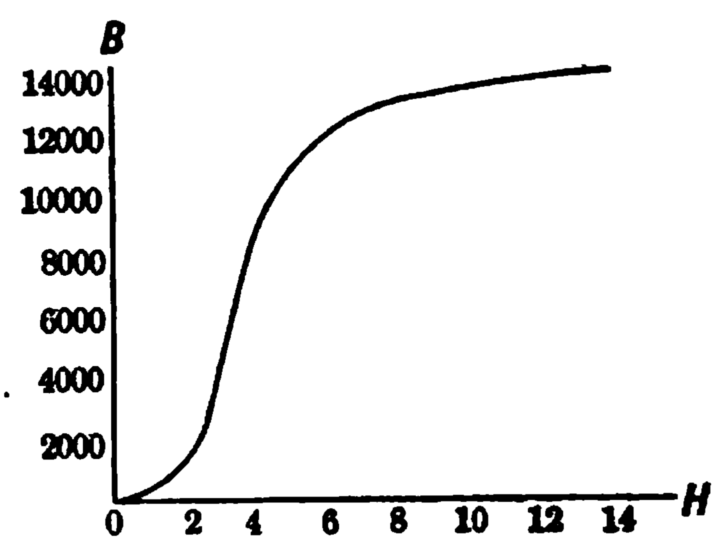


FIG. 358.

BH curves can be made by the magnetometric method as described above (§486), but are as often made by an electromagnetic induction method as described later (§516).

490. Magnetic Permeability.—We have seen that the effect of placing a piece of iron in a magnetic field is to increase the

magnetic lines from H per square centimeter in air, to B lines per square centimeter in iron. Faraday expressed this fact by saying that the iron has a greater *conductivity for magnetic lines* than air has. The same fact is now more frequently expressed by the term *magnetic permeability.* The magnetic permeability, μ, of a substance is defined as *the ratio of the induction to the magnetizing force in the substance,* or

$$\mu = B/H$$

The table below shows the permeability of a specimen of iron for various values of H. It is seen that beyond the bend of the curve, where the magnetization approaches saturation, the permeability decreases rapidly. Knowing the permeability of a certain kind of iron for a given induction, we can find the magnetizing force required to produce the induction. The magnetic susceptibility k, and the permeability are connected by the relation $\mu = 1+4\pi k$. This follows directly by substituting values for B and I in the expression $B = H+4\pi I$, and then dividing by H.

H	I	$k=\frac{I}{H}$	B	$\mu=\frac{B}{H}$
0	0	..	0	....
0.32	3	9	40	120
0.84	13	15	170	200
1.37	33	24	420	310
2.14	93	43	1,170	550
2.67	295	110	3,710	1,390
3.24	581	179	7,300	2,250
3.89	793	204	9,970	2,560
4.50	926	206	11,640	2,590
5.17	1,009	195	12,680	2,450
6.20	1,086	175	13,640	2,200
7.94	1,155	145	14,510	1,830
9.79	1,192	122	14,980	1,530
11.57	1,212	105	15,230	1,320
15.06	1,238	82	15,570	1,030
19.76	1,255	64	15,780	800
21.70	1,262	58	15,870	730

491. Magnetic Circuit.—Maxwell's term magnetic flux or flow of induction has lead to the development of the very useful conception of a "magnetic

circuit" analogous to that of an electric circuit. In an electric circuit we have an electric flux; that is, an electric current i produced by an electromotive force E, in a circuit of conductance C or of resistance R, where $R = \frac{i}{C}$ and the relation $i = CE = \frac{E}{R}$ holds (see Ohm's law, §442). In a magnetic circuit a magnetic flux of N is produced by a magnetising field (*i.e.* "a magnetomotive force") around a circuit of iron or other materials of greater or less magnetic permeability. Let us take the simple case of an iron anchor ring wound with a coil of n turns. From §430, we have $H = \frac{4\pi ni}{L}$, where L is the length of the mean length of the core of a circular solenoid. Then the induction $B = \mu H$, where μ is the permeability of the core. If S is the area of the cross-section of the core, we get the total flux $N = BS = H\mu S$. Substituting the value of H, we get $N = \frac{4\pi ni\mu S}{L}$. This can also be written in the form $N = \frac{4\pi ni}{\left(\frac{L}{\mu S}\right)} = \frac{\text{magnetomotive force}}{\text{magnetic resistance}}$, if we let $4\pi ni =$ m.m.f. the magnetic motive force, and $\frac{L}{\mu S}$ equal the magnetic "resistance" or magnetic reluctance. The magnetic resistance of a circuit thus varies directly as the length of the circuit, inversely as the cross-section, and as some constant $\left(\frac{1}{\mu}\right)$ which we may call the specific magnetic resistance of the material.

FIG. 359.

This parallelism to the electric circuit can be extended to a composite circuit, where there are two air gaps as in the case of the magnetic field of a dynamo, and in such cases becomes of great convenience in calculations.

492. Effects of High Permeability. Magnetic Shielding.—When a magnetic substance is placed in a magnetic field, it changes the distribution of lines of force. This is shown by the arrangement of iron filings about an iron disk placed in a uniform field (Fig. 359). The lines tend to go through the iron, because of its higher permeability. Fig. 360 taken from a paper of Lord Kelvin, shows the field around and through a sphere of high permeability.

Fig. 234e (§368) shows the lines about a thick iron ring in a

uniform magnetic field. It is to be noticed that the lines are diverted toward the ring on account of its higher permeability; and also that the filings on the inside of the ring show little if any directive action. This "screening" effect of soft iron is made use of in protecting sensitive galvanometers from magnetic disturbances (§437). The same principle is used in shielding the interior coils of the ring armature (§519).

493. Magnetization and Temperature.—When soft iron is raised to a temperature of 785°C., it ceases to be ferro-magnetic, though remaining slightly magnetic. This temperature is called the critical temperature. For nickel, the critical temperature is 340° C., for cobalt, 1075°C., for magnetite 535°C., and for a certain hard steel Hopkinson found a value of 690°C. The permeability of a substance for strong magnetizing forces in general decreases as the temperature rises, though there are some variations from this that are not yet explained.

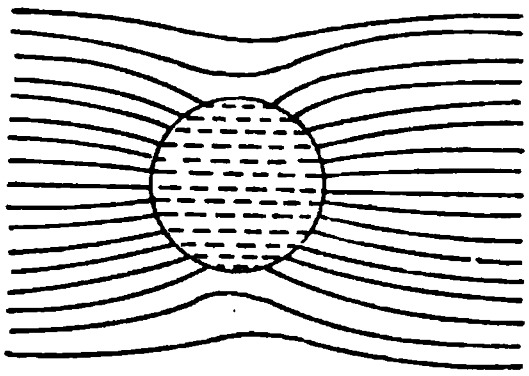


FIG. 360.

494. Diamagnetic Substances.—A bar of bismuth, suspended between the pointed poles of a strong electromagnet takes a position at right angles to the lines of force, and a small ball of bismuth is repelled from strong to weak parts of the field. This action is just opposite to that of iron in a magnetic field, and hence Faraday called bismuth and other bodies showing similar magnetic action, *diamagnetic* substances. Bismuth, antimony, copper, zinc, silver, lead, glass, etc., are diamagnetic. The action of bismuth, the strongest of the diamagnetic substances is however feeble compared to the magnetic action of iron, nickel and cobalt.

495. Para-magnetic and Ferro-magnetic Substances.—Faraday found that many bodies, supposed to be non-magnetic show magnetic properties in the field of a strong electromagnet. Such bodies are called *para-magnetic* substances. The strongly para-magnetic substances, iron, nickel, cobalt, Heusler alloy, are known as *ferro-magnetic* substances. The following table gives the susceptibilities of some substances, negative susceptibilities indicating diamagnetic substances. The susceptibility of the ferro-magnetic substances as we have seen depends upon the

magnetizing force. As the susceptibility is changed by slight impurities, the values given by different observers vary somewhat. Since the permeability μ is equal to $1+4\pi k$, it is seen that negative susceptibilities indicate permeabilities less than unity.

Iron-silicon alloy melted in a vacuum.........	40,000+	Aluminum.....	0.0000018
Ordinary soft iron, max.	200+	Air............	0.000000032
Nickel, max............	23+	Copper........	−0.00000082
Cobalt, max............	13.8+	Lead..........	−0.00000124
Oxygen at 182° C......	0.000324	Silver.........	−0.00000151
Platinum..............	0.000029	Antimony......	−0.0000052
		Bismuth.......	−0.0000138

496. Ampere's Theory of Magnetism. Electron Theory of Magnetism.—The "molecular" theory of magnetism (§366) explains the structure of a magnet but shifts the problem of the nature of "magnetism" to the elementary magnets. Ampere suggested that the molecule of a magnetic substance is a magnet because it has an electric current flowing about it; as we have already seen (§427) a circuit in which a current flows has N and S polarity like a magnet. "According to Ampere's theory," says Maxwell, "all the phenomena of magnetism are due to electric currents, and if we could make observations of the magnetic force in the interior of a magnetic molecule, we should find that it obeyed exactly the same laws as the force in a region surrounded by any other electric circuit." That is, Ampere's theory makes magnetism a section of electrokinetics.

How the elementary electric current started and how it can continue to flow about the "molecule" without consuming energy, were not explained by Ampere. In the *electron theory of magnetism* which is an extension of Ampere's theory, the electric current is explained as due to electrons or corpuscles which revolve about the atom. The magnetization of the elementary magnet is by this theory due to the magnetic action of the revolving electrons.

According to the electron theory of matter, one or more electrons are revolving about every atom. If the number of electrons revolving clockwise is equal to the number revolving anti-clockwise, then the atom is

non-magnetic. But if such an atom is brought into a magnetic field, the effect is the same as bringing a closed circuit into the field, and in accordance with Lenz's law (§501), magnetic lines are set up which are opposite to those of the field and repulsion results. The atom is thus diamagnetic. The character of the electronic orbit is changed in this case by addition of another motion. If the number of electrons having one direction of rotation is greater than the number having the opposite direction, then the atom is naturally para-magnetic. In the case of para-magnetic atoms, it is usually assumed that not only are all the velocities in the electronic orbits changed, but also the planes of the orbits and probably the atoms themselves are rotated by the external field.

While the electron theory of magnetism is the most promising of present theories, it is not developed in a form to explain all the facts.

The discovery by Heusler of a strongly ferro-magnetic alloy of manganese, copper and aluminum, the component parts of which are practically non-magnetic, has introduced new questions as to the nature of magnetism. As these ferro-magnetic alloys all contain manganese, or the related element chromium, it seems probable that manganese and chromium become ferro-magnetic under certain conditions of temperature and combination.

497. Residual Magnetic Effects. Hysteresis.—It is well known that hardened steel retains its magnetization, not only after the magnetizing force is removed, but also against opposing magnetic fields of considerable strength. Even soft iron shows small magnetic effects of the same kind. These after-effects are due to the resistance or "reluctance" which the molecular magnets experience to free rotation. To overcome this resistance to changes of magnetization, work is required. In the case of iron subjected to thousands of magnetic reversals in a minute, as in the alternating current transformer, the energy expended in the cyclic changes of magnetization may become a large quantity.

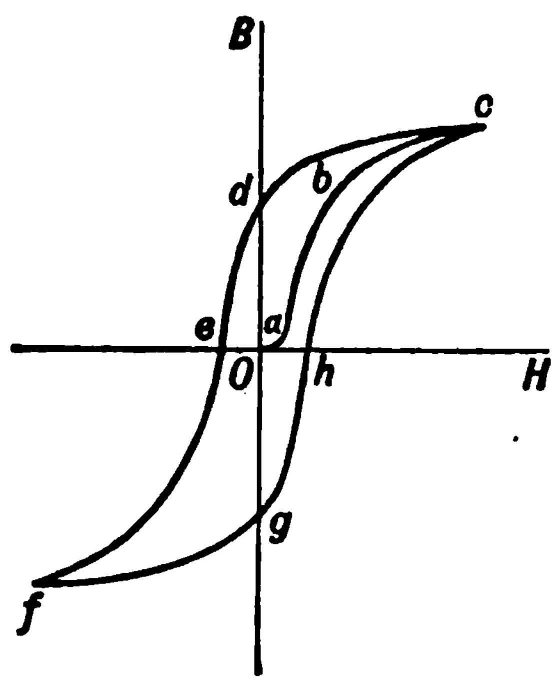


FIG. 361.

To study and measure these phenomena of magnetic lag, the magnetization curve for a complete cycle is made. Starting with non-magnetized iron, it is magnetized by an increasing magnetic field until it approaches saturation. This gives us the curve *Oabc* sometimes called the "virgin curve." Then the magnetizing field is gradually reduced to zero, giving the curve *cd*, which shows that for zero magnetizing field *H*, there

is remanent magnetism represented by Od. This gives a measure of the retentivity or "remanence." The magnetizing force H is now reversed, and for a value $-H = Oe$, the iron is demagnetized. This value of $-H$ measures the *coercive force* of the meal. As $-H$ is increased to a maximum, decreased, reversed, and $+H$ increased in the original direction, the parts of the curve through f, g and h are traced, until the loop is completed at c. There is thus shown to be a lagging of the magnetization behind the magnetizing force during the cycle.

In the above process the elementary magnets have undergone a double reversal in direction, and work has been done against what is sometimes called "magnetic friction." Ewing has shown by his model (§366) that this "magnetic friction" can be explained by the mutual actions of the elementary magnets. The phenomenon is called *hysteresis*. The hysteresis loss is stated in ergs per cu. cm. per cycle and is proportional to the area enclosed by the hysteresis loop. This energy is dissipated in heating the iron. It should be noted that this is entirely different from the heating due to induced eddy currents in the iron (§507).

Iron varies greatly in its quality as regards hysteresis. Thus Ewing found for a certain soft iron at a maximum induction of 5000 lines, a loss of 910 ergs per cu. cm. per cycle and at an induction of 9000 a loss of 2300 ergs per cycle. For cast iron at the same maximum induction the loss may reach a value of ten times the above. It may be added that a loss of 2300 ergs per cu. cm. per cycle represents 1.36 watts per pound of iron per 100 cycles per second, or over three and a half H. P. per ton of iron.

498. Energy of Magnetic Field.—A magnetic field contains energy in the form of strains of some kind in the "ether." The amount of this energy per cubic centimeter is calculated as in the case of an electrostatic field (§419), and is given by the similar expression energy $E = \mu H^2/8\pi$, where E is in ergs, H is the intensity of the field and μ is the permeability of the medium. The importance of this energy of the magnetic field appears in electromagnetic induction. In the case of the alternating transformer (§520), the energy of the primary circuit is converted into the energy of the magnetic field and this magnetic energy is then transformed into the electrical energy of the secondary circuit. In the impedance or choking coil (§523), the energy of the circuit is transformed into the energy of the magnetic field, and then transformed back into the energy of the electric circuit as many times a second as there are alternations.

ELECTROMAGNETIC INDUCTION

499. Induced Electric Currents.—On November 24, 1831, Michael Faraday described to the Royal Society of London a series of experiments showing that electric currents can be produced in a closed conducting circuit, (*a*) by moving neighboring magnets; or (*b*) by changing the current in a neighboring electric circuit; or (*c*) by moving a neighboring electric circuit. An electric current thus produced is said to be *induced*, and the phenomenon is called *electromagnetic induction*. Few discoveries in science have had such important practical results as this discovery of Faraday's. Almost every modern industrial application of electricity depends upon electromagnetic induction.[1]

500. Faraday's Experiments.—The experiments on induced currents made by Faraday were the following: (I) A coil of wire *B* forms a closed circuit through a sensitive galvanometer *G* (Fig. 362). When the pole of a magnet is brought up to *B*, a momentary current is induced in *B*, and the galvanometer needle is deflected. When the magnet pole is removed, a momentary current is again induced, but in the opposite direction to that upon approach. The following facts may be noted: (*a*) The essential motion is relative, that is, moving the coil to or from the magnet produces the same effect as moving the magnet; (*b*) the current lasts only during the time of motion; when the magnet and the coil are relatively at rest, there is no induced current; (*c*) bringing up a *N* pole to a coil induces a current anticlockwise as seen from the pole; that is, the induced current makes this face a *N* face (§427). Thus the approaching *N* pole is repelled by the magnetic action of the induced current. Removing the *N* pole induces a clockwise current, that is, makes the coil face a *S* face. Thus the *N*

FIG. 362.

[1] Working at the same time, Joseph Henry discovered independently the fundamental facts of electromagnetic induction, and probably even anticipated Faraday in some cases. But Professor Henry worked under many disadvantages in the isolated town of Albany, New York, and his discoveries were not widely known at the time they were made.

pole is attracted as it is removed. An approaching S pole induces a current in the same direction as a receding N pole, and *vice versa* (Fig. 363). Or in general, **the magnetic action of the induced current opposes the motion of the magnet.** This is evidently a case of action and reaction. If the approaching magnet were attracted by the induced current, it would require no work to bring the magnet up, and we would get an electric current, which represents energy, without the expenditure of work. This would be contrary to the principle of the conservation of energy (§341).

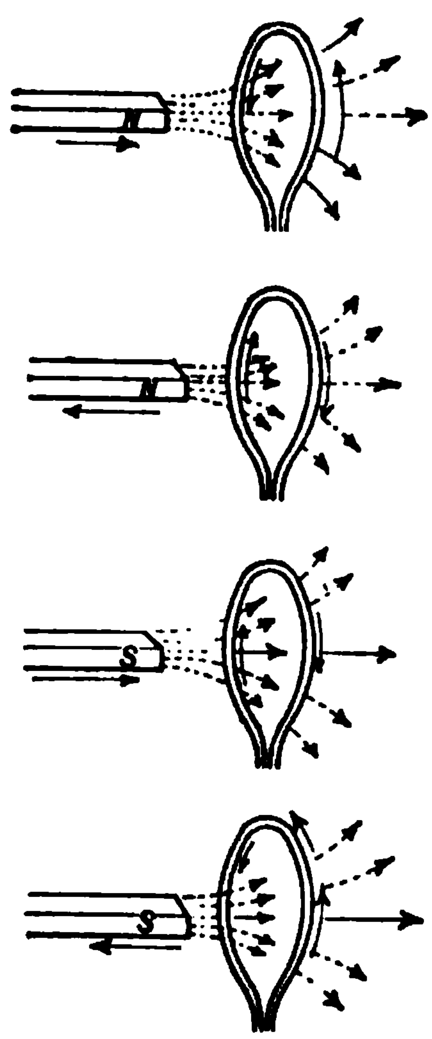


FIG. 363.

We can now describe the above experiment in the convenient terms of the magnetic field and the magnetic lines of force, as conceived by Faraday (§368). The magnet is surrounded by a magnetic field, and the lines of force emerge from the N pole, and enter at the S pole. The motion of the magnet thus changes the number of lines of force included by the coil. The experiment thus shows that, (*a*) a change of the number of lines of magnetic force included by a circuit induces a current in the circuit; (*b*) the current induced is proportional to the rate of change of included lines of force; and (*c*) the magnetic lines from the induced current increase as the magnetic lines from the magnet decrease through the circuit, and *vice versa.* The positive direction of change of lines is to be reckoned in the same direction for both magnet and current. Faraday's other experiments are now easily described.

(II) Substitute for the magnet NS, **a coil carrying an electric current.** A is thus surrounded by a magnetic field (§427), and moving A in front of B, changes the number of magnetic lines included by B, and thus induces an electric current in B during the time of motion. When A is approaching B, the opposing faces of the two coils are either both N or both S; and for this case the induced current in B must be *inverse* in direction to the current in A. Similarly it is seen that upon the receding of A, the induced current is *direct* to that in A.

The coil *A* carrying the original or inducing current, is called the primary coil (*Pr*), and its current the primary current. The coil *B* is called the secondary coil (*Sc*), and the induced current, the secondary current. A current in the same direction as the primary current is called *direct*, and a current in the opposite direction is called *inverse.*

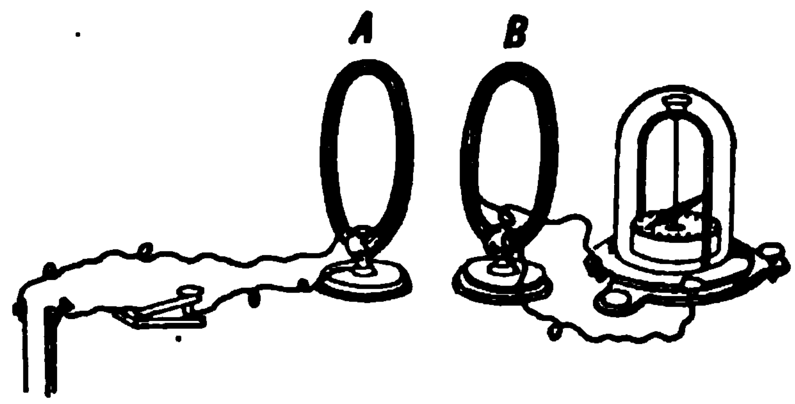


FIG. 364.

(III) With two coils *A* and *B* as before, we can change the number of lines of force through *B*, by changing the current in *A* (Fig. 364). Thus we find that, *making* or *increasing* the *Pr* current induces a momentary inverse current *Sc* in *B*; and that, *breaking* or *decreasing* the *Pr* current induces a *direct* current in the secondary circuit.

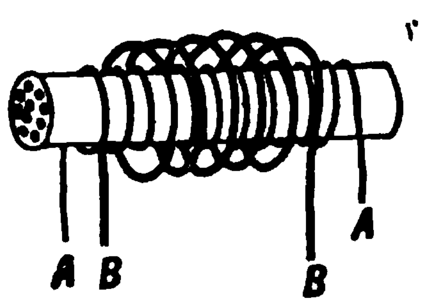


FIG. 365.

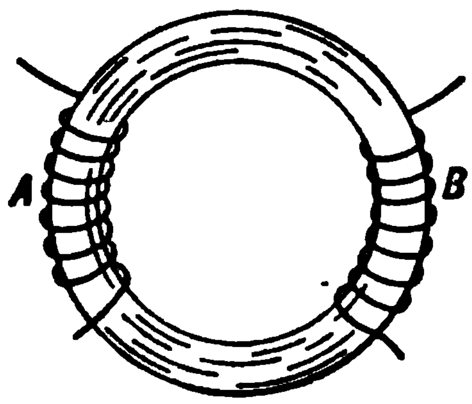


FIG 366.

The coils *A* and *B* may be placed anywhere, provided that the magnetic lines from *A* pass through *B*. Coils wound over or alongside of each other on a cylinder are common arrangements for obtaining maximum induced currents (Fig. 365). Two straight parallel wires may similarly form primary and secondary circuits. This is often the case in telephone circuits, and is the cause of the "cross talking" in the lines.

(IV) If *A* and *B*, one or both, have an iron core, the induced currents are greater, but in the same direction as without the iron cores. This is easily explained in terms of the magnetic lines. Iron has a greater magnetic permeability (§491) than air, so that a given change in the primary current produces a greater change in the magnetic flux through the secondary circuit, and thus causes a greater induced current. Figs. 365 and 366

how common arrangements of the primary and secondary coils on iron cores. The arrangement shown in Fig. 366 is one Faraday used in his earliest experiments.

501. Lenz's Law.—In 1834, Lenz stated the following important relation between the induced current and the motion of the electrical circuit or the magnet causing the induction: *The induced current is in such a direction as to oppose by its electromagnetic action the motion of the magnet or the coil which produces the induction.* We have already seen that this holds in the case of the motion of a magnet, and it is easily seen that it also holds for two coils moving relatively to each other.

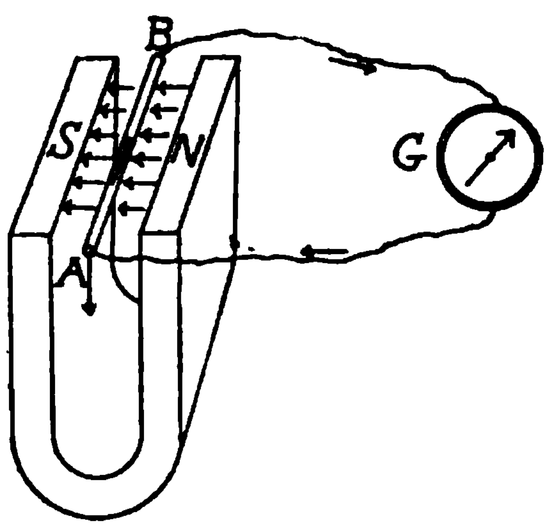


FIG 367a.

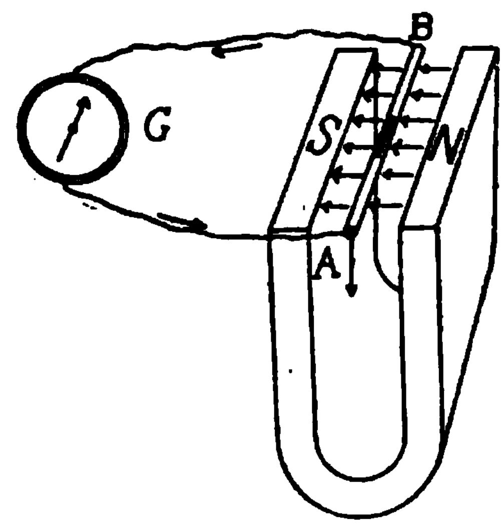


FIG. 367b.

Lenz's law can be extended to the case of a current induced by the variation of a primary current, but the reactions are then purely electromagnetic. When the current is induced in a secondary coil B by making or breaking the current in a primary coil A, we have a reaction of B on the current in A. That is the electromagnetic induction is mutual. We thus have the following series of actions and reactions: Starting a current in A produces magnetic lines through A, and part of them pass through B. There is thus an inverse current induced in B. But this induced current started in B produces lines which are opposite to those produced by the primary current in A. The effect of the induced current is thus to oppose and retard the building up of the magnetic field through the coils.

Upon breaking the primary current, the current induced in B is direct; that is, this induced current produces magnetic lines which are in the same direction as those produced by A. The

effect of the induced current in B is thus to maintain the field, that is, to delay the decrease of the number of lines of force through the coils.

It can thus be seen that in general, the magnetic action of the induced current opposes the magnetic action of the inducing current.

The study of the actions and reactions of the primary and secondary currents, with their energy relations, is very important in understanding the complete theory of the induction coil (§511), and of the alternating current transformer (§520).

502. Induction by Cutting Lines of Force.—One of Faraday's early observations was that "single wires, approximated in certain directions toward the magnetic pole (of a large electromagnet), had currents induced in them." It is often convenient to consider the induction as due to the motion of a single wire across lines of force, or as we often say, due to "cutting lines of force."

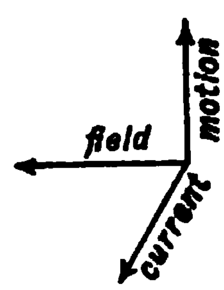


FIG. 368*a*.

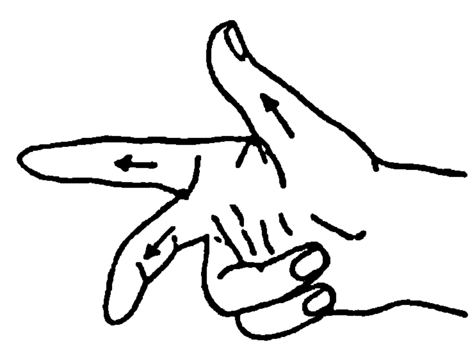


FIG. 368*b*.

Thus suppose the wire AB moved across the field between the poles N and S (*e.g.*, of a U-shaped magnet). In case (*a*) the wire forms part of a complete circuit ABG. In moving AB down, the number of lines through the circuit is increased, and an anti-clockwise current (seen from below) is induced, that is from A to B in the part AB of the circuit. In case (*b*), the wire forms part of the circuit ABG', and the motion downward induces a current in the circuit, so that the current flows from A to B. The current is clockwise as seen from below. When the motion is upward, the current is evidently from B to A. The three directions, of magnetic field, motion, and induced current, are thus mutually at right angles, as indicated in the rectangular axes of Fig. 368*a*. Professor J. A. Fleming has given a convenient rule for remembering these relative directions. Holding the *right* hand as indicated in Fig. 368*b*, with the thumb, the forefinger and the center finger, making right angles with each other, then if the forefinger is held in the direction of the magnetic

field, and the thumb in the direction of the motion, the center finger will indicate the direction of the current.

503. Numerical Calculation of Induced E.M.F.—By Ohm's law, the induced current varies directly as the induced electromotive force and inversely as the resistance of the circuit. The resistance is a constant for the circuit (§442). Experiments show that the electromotive force induced in a circuit is proportional to the rate of variation of lines of force through the circuit, that is,

$$E = -K\frac{N_2 - N_1}{t}$$

where t is the time in seconds, and N_1 and N_2 are the number of lines at the beginning, and at the end of the time interval t. If the variation of N is uniform during the time t, then the e.m.f. induced is constant. When the variation is not uniform the e.m.f. at any instant is given by the differential coefficient, that is, $E = -KdN/dt$. In these expressions K is a constant. If E, N and t are expressed in c. g. s. units, then it can be shown that K becomes unity (§504). To express E in volts, we divide by 10^8 (§441), that is,

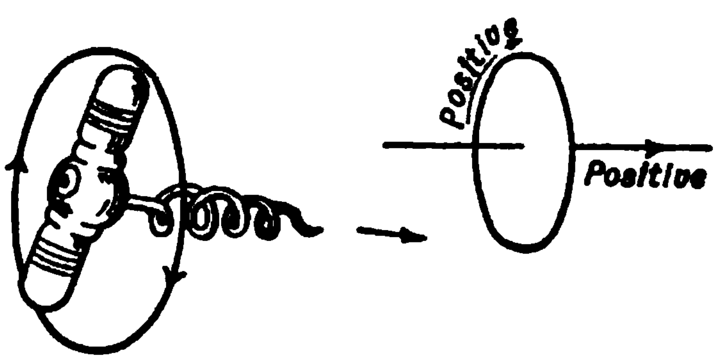


FIG. 369.

$$E\text{ (volts)} = -\frac{N_2 - N_1}{10^8 t} = -\frac{1}{10^8}\frac{dN}{dt}$$

The negative sign is explained as follows: The positive direction of the lines of force is taken as that of the advance or thrust of a right-handed screw, Fig. 369, where the rotation of the same gives the positive direction of the e.m.f., or current. Thus a positive increase of N corresponds to a negative induced e.m.f.

504. A Second Numerical Statement of Induced E.M.F.—It is often convenient to calculate the induced e.m.f. in terms of the number of lines of force cut by a conductor per second. Let H = the strength of the magnetic field, = the number of lines of force per square centimeter section of the field (§374), and let l = the length of the conductor in centimeters, and v = the velocity in centimeters per second. If the motion is perpendicular to the lines of force and also to the length direction of the conductor, then the number of lines cut per second is lvH, or the induced e.m.f. is

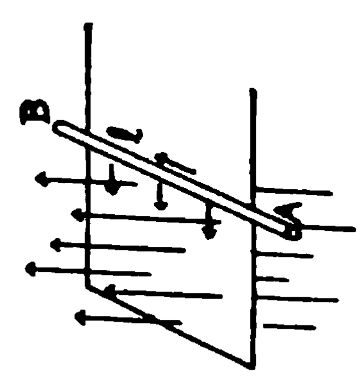


FIG. 370.

$$E = lvH, \text{ or } E\text{(volts)} = \frac{lvH}{10^8}$$

In case the velocity v and the conductor length are not at right angles to the field H, their components at right angles to H are to be taken.

We can derive the above equation by equating the mechanical work done in moving the conductors across the magnetic field to the electrical energy

of the induced current. If i is the induced current, the electrical energy produced in time t is $W=Eit$ (§441). In §529 it is proved that a force $F=Hil$ acts on the conductor, and that this force is in the opposite direction to the motion which induces the current. The distance moved by the conductor in the time t is $d=vt$. Hence the mechanical work is $W=Fd=Hilvt$. Equating the electrical energy to this, we get $E=Hvl$. Hence the constant K in the equation of §503 must be unity for the absolute c.g.s. units.

505. Calculations for Current and Electric Quantity.—In the above section (§503) it has been shown that the induced e.m.f.

$$E=-\frac{N_2-N_1}{t}=-\frac{dN}{dt}$$

The current is then

$$I=\frac{E}{R}=-\frac{N_2-N_1}{Rt}=-\frac{dN}{Rdt}$$

Thus we have $It=-(N_2-N_1)/R$ and $Idt=-dN/R$. But $It=Q$, the total flow of electric quantity in the time t, and $Idt=dQ$, the electric quantity in the time dt. Thus the total quantity of electricity induced is

$$Q=-\frac{N_2-N_1}{R}=-\int\frac{dN}{R}$$

This quantity can be measured by the throw of a ballistic galvanometer, when the time of induction is a small fraction of the period of the galvanometer needle (§439). The above relations are in *absolute units*.

506. Faraday's Disk Dynamo.—One of Faraday's earliest experiments in electromagnetic induction was to rotate a copper disk between the poles of a magnet, the plane of the disk being perpendicular to the field (Fig. 371). A galvanometer circuit was completed by wires sliding on the axle and on the circumference of the disk, and a current was shown by the deflection of the galvanometer during the rotation of the disk. In this machine each radius of the disk cuts the lines of the field at the rate of $\pi r^2 nH$ per second, where πr^2 is the area of the disk, H is the strength of the field assumed uniform, and n is the number of revolutions per second of the disk. Thus

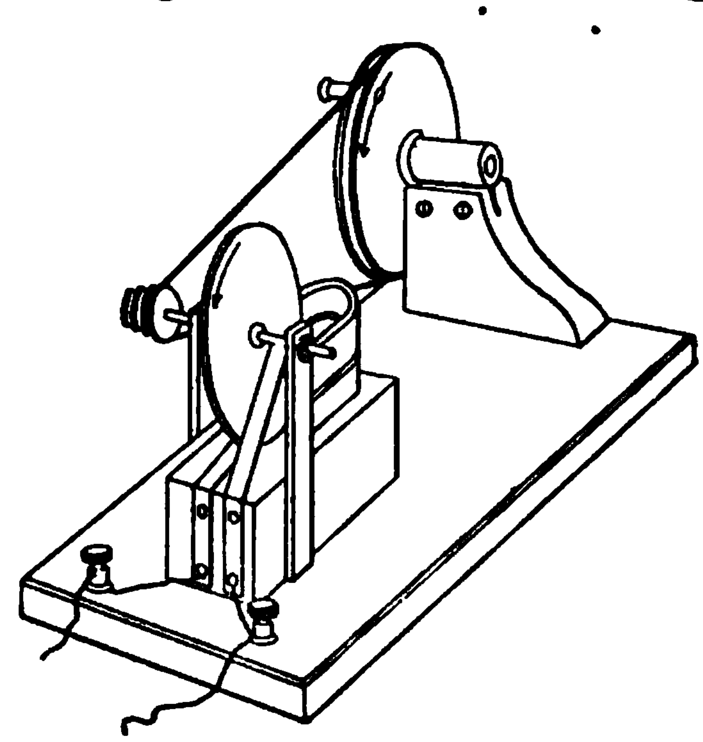


FIG. 371.

$$E\,(\text{volts})=\frac{\pi r^2 nH}{10^8}$$

This arrangement made the first dynamo-electric machine. Forbes and others have attempted to use this as a model for commercial electric generators, but the e.m.f., with any practical diameters and speeds is too small for industrial uses.

507. Foucault or Eddy Currents.—It was observed a number of years before Faraday's discovery of induced currents, that a vibrating magnetic needle quickly came to rest when near or over a copper plate. Arago had in 1824 also shown that a magnetic needle suspended over a rotating copper disk, rotates with the disk (Fig. 372). Both the damping of the needle, and Arago's disk experiment were explained by Faraday as phenomena of electromagnetic induction. The relative motion of the magnet and the disk induces an e.m.f. in the metal disk. The current thus generated circulates in the disk, producing a magnetic action, which by Lenz's law tends to hold the magnet at rest relative to the disk or plate.

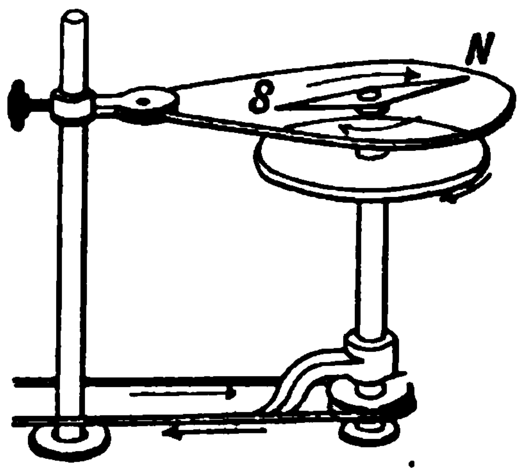


FIG. 372.

Electric currents, thus induced and circulating in a metal mass, are called eddy currents or Foucault currents. The energy of such currents is dissipated in heat. The iron cores of armatures of dynamo machines and transformers are always laminated so as to offer very high resistance to the formation of such currents, and thus to stop the heat losses.

An interesting example of damping by eddy currents is shown in Fig. 373. The pendulum with its copper plate swings freely if the electromagnet is not excited, but is damped immediately when the magnetic field is made.

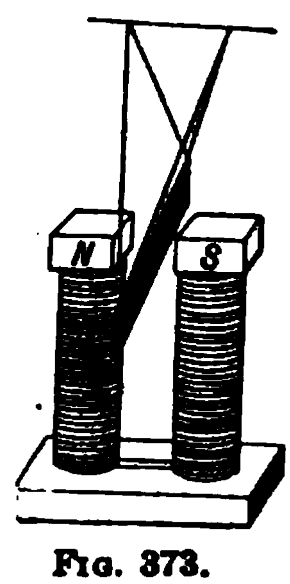


FIG. 373.

This damping action of eddy currents is often taken advantage of in d'Arsonval or movable coil galvanometers (§438), to bring the moving coil to rest quickly. The coil is wound on a closed copper frame in which the eddy currents are generated during the vibration. The coil itself is damped in the same way on closed circuit. The magnetic needles of galvanometers are also often damped by suspending them in openings in copper blocks.

508. Self-induction.—Several persons seem to have observed independently about 1832, that there is a bright spark when a current is broken in a circuit containing an electro-magnet. On making a current in the same circuit, there is no such spark. This extra current at break was noticed by persons receiving a shock upon breaking the circuit of an electromagnet, if they had the terminals in their hands at the time of break. In investigating this, Faraday observed the following facts. Upon breaking the circuit of a helix without an iron core, a similar bright spark is obtained, only less than in the case of an electromagnet. Likewise, when the current in a long straight wire was broken, a spark occurred, only less bright than in the case of the helix. Upon breaking a current in a short wire there was practically no spark. Also in the case of a long wire doubled back on itself, there was no extra spark at break. "The first thought that arises in the mind," wrote Faraday, "is that the electricity circulates with something like momentum, or inertia in the wire, and that thus, a long wire produces effects at the instant the current is stopped, which a short wire cannot produce. Such an explanation is however at once set aside by the fact that the same length of wire produces the same effects in very different degrees, according as it is simply extended, or made into a helix, or forms the circuit of an electromagnet." Faraday then proceeded to show that this extra current was due to electromagnetic induction, from the varying current acting on its own circuit. This phenomenon of a current inducing an extra electromotive force in its own circuit, is called *self-induction*. A circuit includes in general lines of force due to its own current. Breaking the current thus removes the lines of force, or has the same effect as removing a magnet. Suppose the current flows clockwise in a circular circuit; breaking the current then removes positive lines, or is the same as removing a north pole. But this induces a clockwise e.m.f., that is a direct current; this adds itself to the current being broken, and thus causes the bright spark of break. The effect of an iron core is to increase the magnetic induction, and thus to increase the e.m.f. of self-induction at break. Self-induction thus prolongs the current at break, or acts to retard a decrease of the current. When the circuit is wound back on itself, so that it includes no lines of force, there can be no change of lines

of force, and hence no self-induction. Such circuits are said to be non-inductive or inductionless. The wire in resistance boxes is wound doubled from its middle point as is shown in Fig. 321.

When a current is made or increased in an inductive circuit, such as a helix, magnetic lines are put through the circuit. Thus positive lines enter at the face in which the current is clockwise (S face), and this is equivalent to bringing up a N pole. But this induces an anti-clockwise e.m.f., that is an e.m.f. inverse to the starting current. That is, the building up of a current in a coil is accompanied by an induced inverse e. m. f. at the time of the current increase. Here again the self-induction opposes and delays the current changes. Fig. 374 shows the growth and dying away of currents in an inductive circuit as observed in an oscillograph (§524). Helmholtz deduced in 1851 an equation showing the law of the growth of currents in inductive circuits and these oscillograph curves confirm the Helmholtz equation.[1]

FIG. 374.

509. Coefficient of Self-induction or Inductance.—The e.m.f. of self-induction in a circuit thus depends upon the change in the number of lines of force through the circuit, caused by the variation of the current. The number of lines evidently depends upon (*a*) the current I, and (*b*) upon the dimensions of the circuit, and (*c*) the presence of magnetic substances, such as an iron core. For a circuit without iron, N the number of lines included by the circuit varies directly as I, or $N=LI$, where L is the *coefficient of self-induction* or the *inductance* of the circuit. Thus the inductance of a circuit is numerically equal to the increase in the num-

[1] Helmholtz's equation is

$$I=\frac{E}{R}\left(1-e^{-\frac{Rt}{L}}\right)$$

where I, E, R, L and t represent the quantities indicated in this and the next section, and e is the base of the Naperian logarithms. This equation is deduced as follows. The e. m. f. in the circuit at any instant is equal to the impressed e. m. f. less the counter e. m. f. of self-induction, or $e=E-L\,dI/dt$. Hence the current is

$$I=\frac{E}{R}-\frac{L}{R}\frac{dI}{dt}$$

By integration of this differential equation we get the above equation of Helmholtz.

ber of magnetic lines included by the circuit for unit increase of the current. For a circuit with an iron core, this increase of magnetic lines per unit current is not constant, because the magnetic permeability of iron varies with the magnetising force (§491). Hence L the inductance of a circuit with an iron core is a variable depending upon the magnetic curve of the iron.

FIG. 375.

We can also express the inductance of a circuit in terms of the e.m.f. induced for unit rate of change of the current in the circuit. This is shown as follows: The number of magnetic lines through a circuit is $N=LI$; hence the induced e.m.f. is $E=-dN/dt=-L(dI/dt)$. If the rate of change of the current is unity, that is if $dI/dt=1$, then $E=L$. We can thus define unit inductance, as the inductance of a circuit in which unit e.m.f. is induced by unit change of current per second in the circuit. The practical unit of inductance is the *henry* and is equal to 10^9 C.G.S

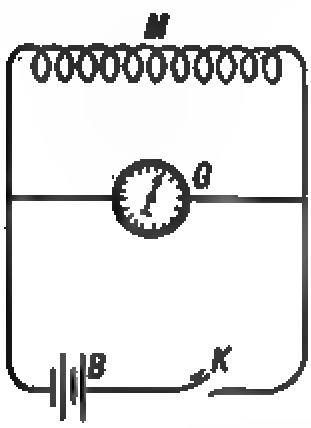


FIG. 376.

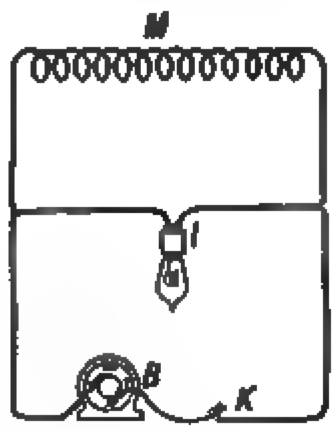


FIG. 377.

units of inductance. The henry can be defined as the inductance of a circuit, in which a change of one ampere per second produces an induced e.m.f. of one volt. Standards of inductance are used in the shape of coils wound on marble or other non-magnetic and permanent cores. These are graduated in multiples or submultiples of the henry. A variable standard of inductance can be made by two coils joined in series and arranged so that they can be rotated in reference to each other, and thus change

the total lines included. Such an inductance standard is illustrated in Fig. 375.

510. Experiments on Self-induction.—To demonstrate extra-currents due to self-induction, Faraday made the following experiment: In a circuit containing a large helix or electromagnet M, there is a galvanometer G, the galvanometer being in parallel with the helix M (Fig. 376). The current from the battery B is made or broken by the key K. In the steady condition, the current divides between the helix and the galvanometer, and there is a steady deflection of the galvanometer needle, say of n degrees to the *right*. A stop is placed so that the needle can not deflect to the right. Upon breaking the current at K, there is a throw of the galvanometer to the *left*, due to the extra-current of break flowing back through the galvanometer circuit. Evidently the extra-current in M is in the same direction as the current being broken.

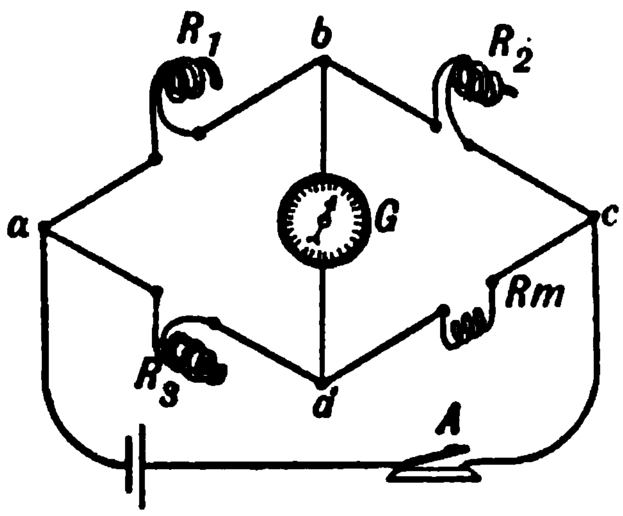


FIG. 378.

A striking variation of the above experiment is to put an incandescent lamp in parallel with an electromagnet, passing just enough current to bring the lamp to a red glow (Fig. 377). Upon breaking the current, the lamp flashes brightly for an instant, due to the e.m.f. of self-induction at break.

The best way of showing the effects of self-induction is by the use of the Wheatstone bridge, as described by Maxwell. In the bridge arrangement, Fig. 378, the resistances R_1, R_2 and R_3 are wound non-inductively (§508), and R_m is the resistance of an electromagnet. The resistances are arranged so that $R_1 : R_2 = R_3 : R_m$. There is accordingly no deflection of the galvanometer G when the current is in a steady state. But upon making the current by closing the key K, there is a momentary throw of the galvanometer needle, the deflection of the needle being again zero when the current reaches a steady state. Upon breaking the current, there is a throw of the needle in the opposite direction to that at make. If all four of the resistances are non-inductive, there are no such momentary throws of the galvanometer needle at make and break. Suppose the current enters at c; then the current reaches its full value in cR_2b sooner than in cR_md, so that there will be a deflection of the galvanometer showing a momentary current from b to d. Upon break the momentary flow will be from d to b.

The explanation of this is evident from §508. The above method of showing the extra-currents has been developed by Maxwell and others into a method of measuring the coefficient of self-induction. For these methods the reader must refer to the laboratory manuals.

511. The Induction Coil.—The induction coil is a piece of apparatus for producing pulsating currents or discharges of high e.m.f. in a secondary circuit, by making and breaking a current

in a primary circuit. The current in the primary circuit may be from a battery with only a few volts e.m.f. In Fig. 379, we have a diagram showing the parts and arrangement of the ordinary induction coil. The primary circuit *Pr* consists of (*a*) a solenoidal coil *P* with a bundle of soft iron wires as core; (*b*) an interrupter *K* for making and breaking the primary current. When the interrupter is mechanical as shown in the figure, there is a condenser joined across the gap to lessen the extra spark of break, and thus cause a quicker break of the current; (*c*) the secondary circuit *Sc* consisting of a solenoidal coil *S* surrounding the coil *P*, and a spark gap *D*. The secondary coil is wound with many turns of fine wire. To increase the insulation, this coil is also wound in disk sections. The primary coil is wound with a comparatively few turns of much larger wire.

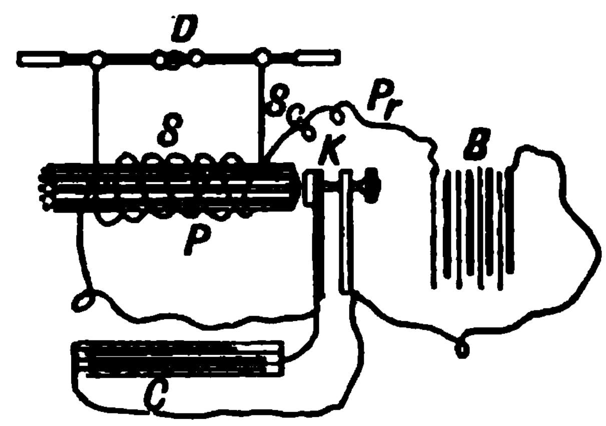


FIG. 379.

Making the primary current produces magnetic lines which thread through the secondary. These lines are removed upon breaking the primary current. Thus there is induced in the coil *S* an inverse e.m.f. at make, and a direct e.m.f. at break of the primary current. The break in most coils is much quicker than the make, and thus the direct induced e.m.f. in *Sc* is so much greater than the inverse induced e.m.f., that the discharge effects are mostly uni-directional. The reason for this is that the growth of the primary current at make is retarded by the inductance of the circuit (§508), while with a good interrupter and proper condenser, the break can be made very sharp. The greater the number of turns of the secondary coil, the greater the induced e.m.f. The resistance of the coil is of course high, and consequently the current small.

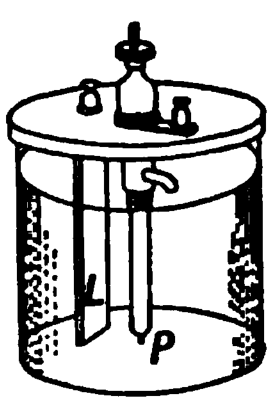


FIG. 380.

In small induction coils the Wagner hammer is the common form of interrupter. This is shown at *K* in Fig. 379, and its action can be easily seen. In large coils, a form often used consists of a brush sliding on a revolving commutator driven by an electric motor. The electrolytic interrupter of Wehnelt is also

frequently used (see Fig. 380). P is a platinum wire in a solution of sulphuric acid, L is a lead plate. Only the point of the wire is exposed to the acid. When P is made the anode, and L the kathode, gas forms at P, interrupts the current, and escapes in bubbles, and thus the current is again made.

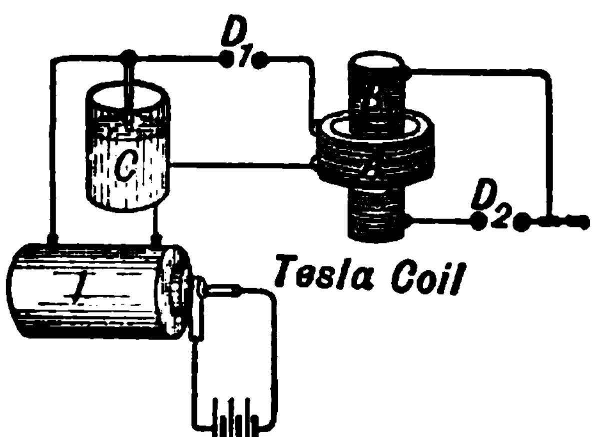


Fig. 381.

512. The Tesla Induction Coil.—To obtain currents of very high frequencies and high electromotive forces, Tesla used a form of induction coil in which the oscillatory discharge of a Leyden jar (§541) is used as interrupter. The terminals of the secondary of an induction coil I (Fig. 381), are connected, one to the inner coating, and one to the outer coating of an insulated Leyden jar C. The circuit is completed through the primary winding of the Tesla coil, and the discharge balls. The primary of the Tesla coil A, consists of a half dozen turns of wires on a non-magnetic core. The coils A and B are separated by air or oil as insulation. The alternations at D from the Leyden jar may have a frequency of several millions per second (§541). Hence the currents induced in B are not only of high e.m.f. but also of very high frequency.

513. Electromotive Force in a Coil Rotating in a Magnetic Field.—In the earth inductor and in many dynamo-electric machines, electric currents are produced by rotating coils of wire in a magnetic field. We take the simple case of a rectangular coil in a uniform magnetic field. The coil $ABCD$ is rotated n times per second about an axis OO', which bisects the coil and is perpendicular to the field (Fig. 382 a, b). Let $l = AB = DC$, and $r = AO = DO$. The direction of rotation looking on the end AOD is anti-clockwise. Let XOX' and YOY' represent the planes through O, respectively parallel and at right angles to the magnetic field. The side AB is evidently cutting lines of force most rapidly at X and X′, where it is moving at right angles to the lines; while at Y and Y' it is moving parallel to the field, so it is cutting no lines, and hence has no e.m.f. induced in it at this instant. The e.m.f. induced in AB is a maximum at X, and

zero at Y, a negative maximum at X', and zero again at Y'. Thus the e.m.f. in AB is from A to B while the coil is moving from Y through X to Y'; it is from B to A while moving from Y' through X' to Y. (These directions follow in accordance with Fleming's rule, §502.) The e.m.f. induced in the opposite side DC evidently adds itself to that in AB to produce a single alternating current in the circuit $ABCD$. The ends BC and DA do not cut lines, and hence have no e.m.f. induced in them. There is thus induced in the coil $ABCD$ an e.m.f. which goes through a complete cycle once in a revolution. It can be shown that this e.m.f. at any instant is proportional to the sine of the angle θ which the plane of the coil makes at

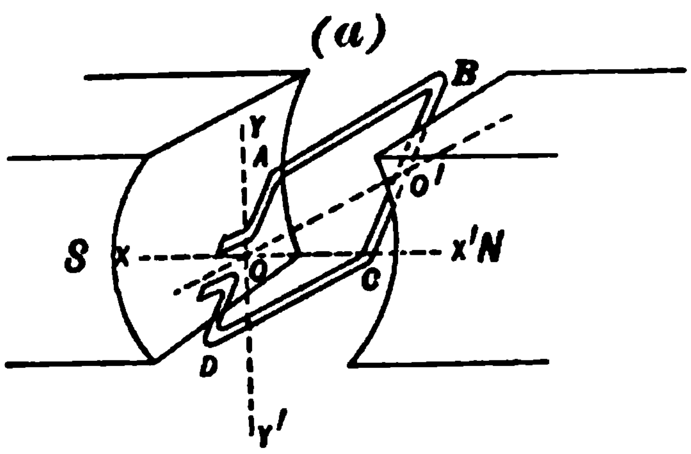


FIG. 382a.

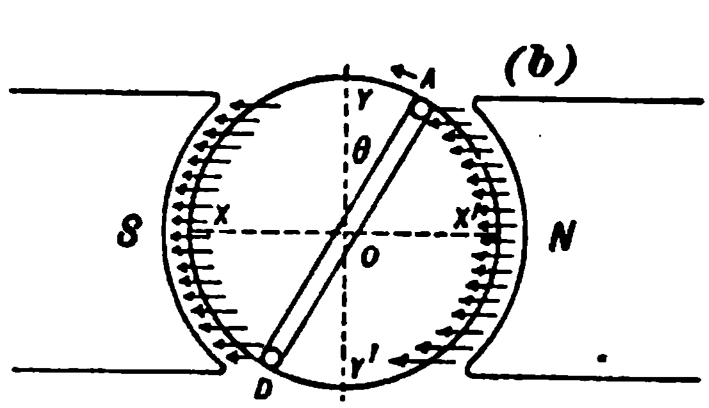


FIG. 382b.

that instant with the plane perpendicular to the field. Thus in Fig. 382b, the e.m.f. is equal to $e = E \sin \theta$, where E is the maximum e.m.f. induced, and θ is the angle of the coil with the plane YOY'.

514. Formula for the E. M. F. in a Rotating Coil.—Let $V=$ the uniform tangential speed of AB (and CD). At the instant when the angle between the coil and the plane perpendicular to the field is θ, this velocity is represented by AR, Fig. 383. The velocity component at right angles to the field is $RS = V \sin \theta$. Let $H =$ the strength of the field ($=$ the number of magnetic lines per square centimeter); then $VlH \sin \theta$ is the number of magnetic lines cut by AB ($=l$) per second. Hence the e.m.f. induced in AB and CD is $e = 2VlH \sin \theta = E \sin \theta$. Here $E = 2VlH =$ e.m.f. when the coil is passing through the points X and X' where it is cutting the lines at the maximum rate, or when $\theta = 90°$ or $= 270°$. The curve, Fig. 384, represents this e.m.f. during a single rotation. The ordinates are proportional to the e.m.f.

and the abscissas are proportional to the angles. Since the rotation is uniform, the abscissas are also proportional to the time. Ordinates above the line represent electromotive forces in one direction, and ordinates below represent electromotive forces in the reverse direction. That is, a rectangular coil rotated uniformly in a uniform magnetic field, has induced in it an alternating e.m.f. which varies as the sine of an angle, and is represented by a sinusoidal curve.

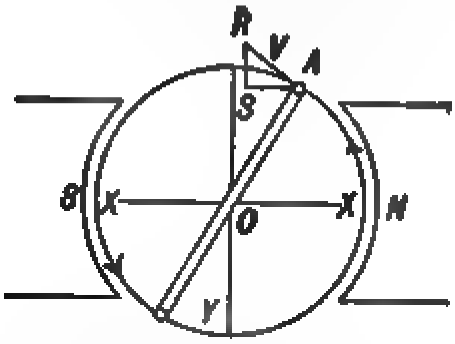


FIG. 383.

FIG. 384.

The formula $e = 2VlH \sin \theta$, can be changed into the form $e = 2\pi n\, N_0 \sin \theta$, where n is the number of revolutions per second, and N_0 is the total magnetic lines through the coil when it is at right angles to the field. To prove this, put $V = 2\pi nr$, and we get $e = 4\pi nrl\, H \sin \theta$. But $2rl = S$, the face area of the coil, and $SH = N_0$. Hence $e = 2\pi n\, N_0 \sin \theta$. It is easy to show that this equation holds for any shape of the rotating coil.

The varying e.m.f. induced in a rotating coil of any form and of area S can be obtained from the relation $E = -\frac{dN}{dt}$. N is the total number of lines, or flux, of magnetic force through the coil and is evidently equal to the flux through the projection of the coil on the plane YY' or $HS \cos \theta$. Hence $E = HS \sin \theta \frac{d\theta}{dt} = HS\omega \sin \theta$ where ω = the angular velocity = $2\pi n$.

FIG. 385.

515. The Earth Inductor.—The earth inductor is a coil, Fig. 385, usually of a large number of windings, which is mounted so that it can be rotated about either a horizontal or a vertical axis. Suppose the axis vertical and the plane of the coil at right angles to the magnetic meridian. By revolv-

ing the coil through 180°, the magnetic flux is taken out of one face and put in at the other face. Let $H =$ the horizontal intensity of the earth's magnetic field, $S =$ the area of the coil face, and $R =$ the resistance of the circuit. Then the quantity of electricity flowing in the coil during a rotation of 180°, is $q = (2HS)/R$ (§505). This can be measured by the throw of a ballistic galvanometer. Similarly when the axis is placed horizontally, $q' = (2VS)/R$. where V is the vertical component of the earth's field. We thus get $(q'/q) = V/H = \tan\phi$, where ϕ is the dip or inclination (§386).

516. Use of Induced Currents to Compare Fields and to Measure Magnetic Induction.—From the above we see that the intensities of two magnetic fields can be compared by quickly rotating the same coil in the two fields and comparing the throws of a ballistic galvanometer. More often a small coil, often termed a "flip coil" is quickly jerked out of the field, and the throw of the galvanometer noted.

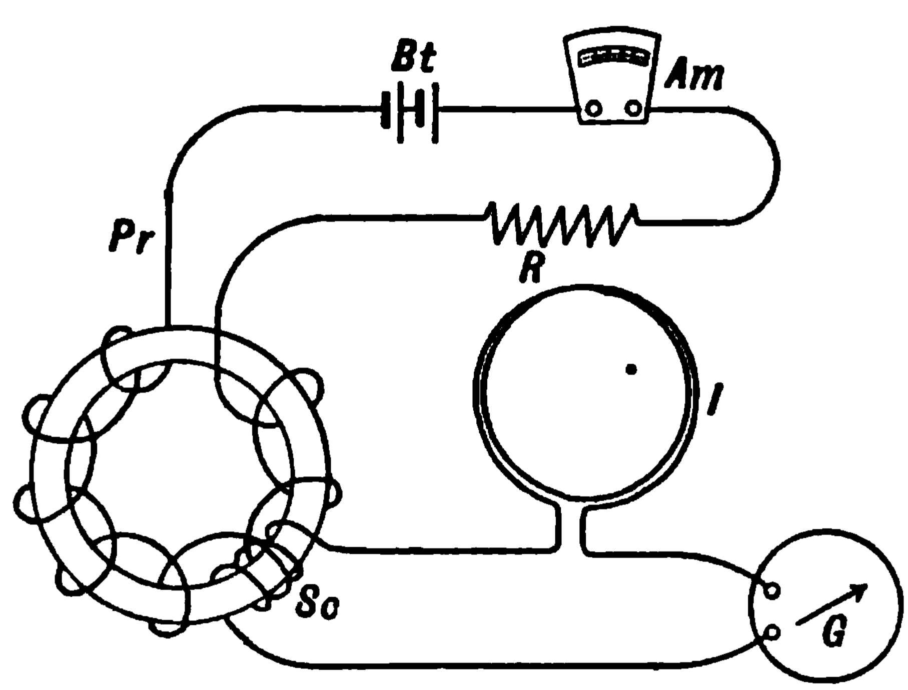


FIG. 386.

The change of magnetic induction B, in a ring magnet is determined in Rowland's method of obtaining BH curves, by observing the galvanometer throw due to the current induced in a secondary coil Sc. The arrangement is shown in Fig. 386. The iron anchor ring is wound uniformly with the magnetizing coil Pr. The current is measured by the ammeter Am, and regulated by the rheostat R. Changes in the current in Pr produce changes in the magnetic induction B through the iron. Starting from zero the current is increased by steps, and the increments in the induction B are calculated from the galvanometer throws. The total induction is gotten by adding the increments. The magnetizing force H is calculated from the dimensions of the coil Pr and the current i. The ballistic galvanometer G is calibrated by observing the throw from revolving the earth inductor I in the earth's magnetic field. The values of H and the corresponding values of B thus determined are then plotted to give the usual curves (§490).

517. Simple Alternating Current Dynamo.—In Fig. 387, we have a coil revolving in the field between the poles of a magnet. The ends of the coil are connected with the insulated metal rings

N and *M*, on the shaft *OO′*. Metal springs or "brushes" rest or slide on these collector rings. Thus the current induced in the coil *ABCD* flows through the external circuit. Such a machine forms a simple alternating current dynamo. By winding the coil on an iron cylinder, the intensity of the magnetic field is increased, and thus a greater e.m.f. is induced. But this iron core being itself a conductor will have eddy currents induced in it, unless it is laminated, so as to make the resistance infinitely great in the direction of the induced e.m.f. Hence the core is built up of insulated iron disks as represented in Fig. 388. Commercial alternating current dynamos are always multipolar. One of the reasons for this is that for practical electric lighting and power transmission, frequencies of from 25 to 125 alterna-

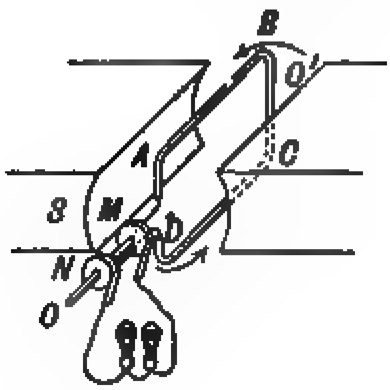


FIG. 387.

FIG. 388.

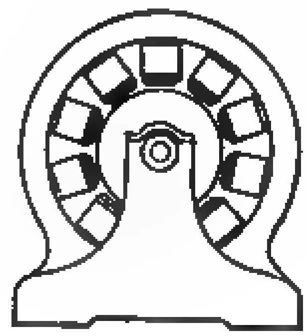


FIG. 389.

tions per second are desirable. A common frequency for most purposes is now 60 alternations per second. To get such frequencies with safe speeds it is necessary to have field magnets with multiple poles. Fig. 389 shows a multipolar alternating current dynamo.

518. Simple Direct Current Dynamo.—To obtain a current constant in direction, a commutator is used instead of collector rings. Fig. 390 represents a two-part commutator. This consists of a copper ring, which has been cut into half rings. These half rings are insulated and form the ends of the coil *ABCD*. The brushes *R* and *S* are set 180° apart, so that one rests on one-half of the commutator, while the other rests on the opposite half. These brushes thus make the connections for the current with the external circuit. The brushes are set so that the connections with the external circuit are reversed, just at the instant in which

the current in the rotating coil is reversed (that is, approximately as the coil passes through the plane perpendicular to the field). The current in the external circuit is thus uni-directional, and varies as represented in Fig. 391. The above is a simple direct current (D. C.) dynamo with a two-part commutator. In Fig. 392 we have two coils at right angles to each other; by joining to a proper commutator we obtain in the external circuit, the current represented in Fig. 393. This is the sum of two pulsating

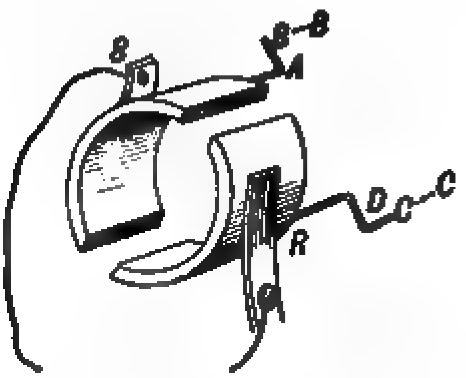


FIG. 390.

FIG. 391.

currents, which differ in phase by 90°. It is seen that the per cent. of variation is much less than in the current from a dynamo with a two-part commutator. In modern direct or continuous current dynamos there are often hundreds of coils, joined to a commutator of many sections, and the resulting current is practically constant. The inductance of the circuit also operates to lessen the variations of the current in these machines.

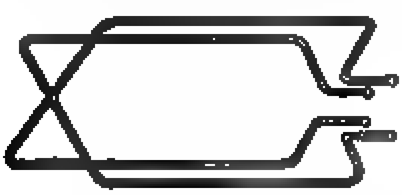


FIG. 392.

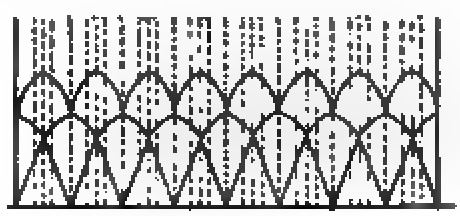


FIG. 393.

519. Dynamo-eléctric Machines.—The parts of a dynamo-electric machine, or a dynamo, are (1) the field magnets for producing the magnetic field; (2) the armature, or the coils in which the currents are induced. The armature coils are nearly always on a laminated iron core, and are supplied with slip rings or a commutator, and brushes to make connection with the external circuit. In the simple forms of dynamos described in

previous sections, the armature revolves, and the field magnets are stationary. In large dynamos for high electromotive forces, the armature is often made the stationary part and the field magnets revolve. In one type of A. C. dynamos, the revolving

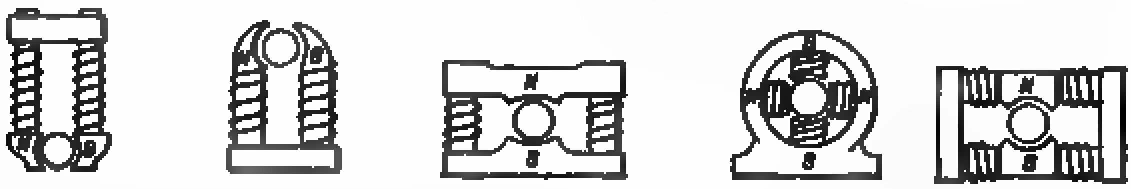


Fig. 394.

parts are iron masses which change the magnetic flux through the armature coils, the magnet coils being also stationary. This type of dynamo is called the "inductor" form.

Dynamos are "direct" current (D. C.), or alternating current (A. C.), according to the character of the e.m.f. at the terminals

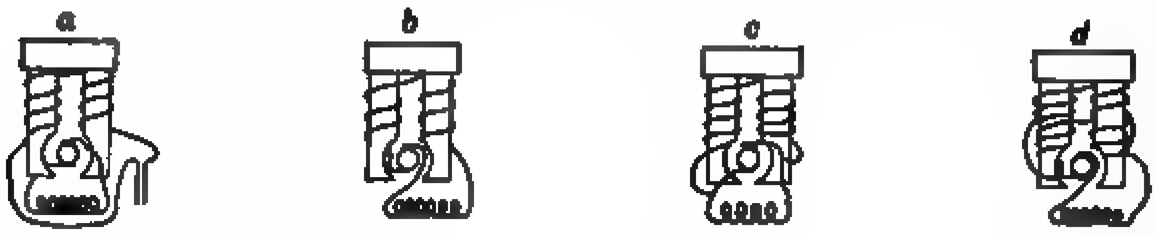


Fig. 395.

of the machine. The field magnets may be bi-polar or multipolar. Several common forms of field magnets are shown in Fig. 394. In some of the early dynamos, permanent magnets were used for field magnets. Such machines were called magneto-electric machines or magnetos. The small machines often used in telephone call boxes are magnetos. But the field magnets of all modern power dynamos are electromagnets.

Fig. 396.

Fig. 395, *a*, *b*, *c*, *d*, shows the different methods in which the field magnets are excited. These are (*a*) separately excited, that is the current in the field coils comes from a separate generator; (*b*) "series wound," that is the field coils are in series

with armature and the external circuit, so that all the current of the armature passes through the field coils as well as the external circuit; (c) "**shunt wound,**" that is, the field coils and the external circuit are in parallel, so that only a part of the current of the armature passes through the field coils; (d) "**compound wound,**" that is, part of the field coils are series and part shunt windings. The choice of windings of the field coils is largely a question of regulation of the electromotive force under different loads. For a discussion of these methods, the student must consult special manuals.

The two most common forms of D. C. armatures are: (a) the **ring armature,** sometimes called the Gramme armature, after its inventor, and (b) the **drum armature.** The ring armature is represented diagrammatically in Fig. 396. The coils are wound around a closed ring of soft iron, and connected as indicated. The core of the ring is laminated. Only the wires on the outside of the ring are inductors, as the wires on the inside are shielded magnetically. The course of the lines of force is indicated in the figure.

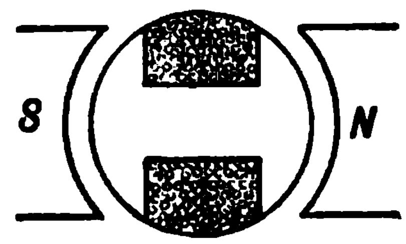


FIG. 397.

The first and the simplest form of drum armature is the "shuttle" armature used by Siemens in his machine of 1856. A section is shown in Fig. 397. The iron core increases the magnetic flux through the coil. This form is now used only in small magnetos. The winding and commutator connections of a modern drum armature are very complicated. The coils may be on the surface or in tunnels or grooves in the core of the armature. The magnetic lines go from pole to pole as indicated in Fig. 398. For the study of the forms of armature and of their actions and reactions, the student must consult special treatises on dynamo-electric machinery.

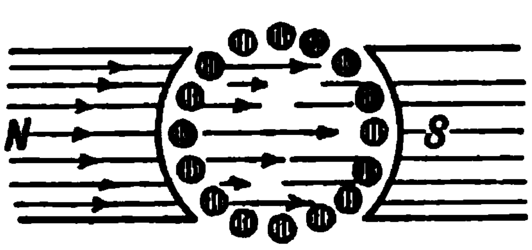


FIG. 398.

520. The Alternating Current Transformer.—The alternating current transformer is a form of induction coil, used for transforming alternating currents of one potential into alternating currents of a different potential. It consists of a primary coil *Pr*, and a secondary coil *Sc*, and a laminated iron core to increase the magnetic flux. It is most commonly used to "step down"

from a higher voltage to a lower voltage. The energy of the secondary current in well-designed transformers, is equal within a small percentage to the energy of the primary circuit. Thus a current of one ampere at 1000 volts is transformed into approximately 10 amperes at 100 volts. The *Pr* coil has in this case ten times the number of turns of the *Sc* coil. The only limit to the potentials that can be obtained with transformers is that of insulation. The coils of tansformers for high potentials are generally immersed in a high insulating mineral oil.

521. Advantages of Alternating Currents in Power Transmission.—Within recent years, electric power has been transmitted scores of miles, and alternating currents are used exclusively on these long distance power lines. The reasons for this general use of the alternating current in transmitting electric energy over longer distances are (*a*) the ease of transforming the alternating current from high to low potentials; (*b*) the possibility of securing high insulation in alternating current machinery; (*c*) the invention of the A. C. induction motor (§538).

Electric power is measured by the product of the current and of the potential, or equals Ie; thus the same power can be transmitted at a high potential with a small current, or at a low potential with a correspondingly larger current. But the weight of copper needed in the lines increases rapidly with the current, since heating effects vary as I^2R (§458). There is thus a great economy in the transmission of electric power at high potentials, and indeed only in this way is it commercially possible. Potentials of from 30,000 to 60,000 volts are in use for such transmission. But these high potentials cannot be used safely in lamps or in moving apparatus, so that it is necessary to transform to lower potentials before using the currents. For alternating currents, this can be easily and efficiently done by the A. C. transformer (§520). Such changes of potentials are not possible with any D. C. apparatus. The use of high potentials with alternating currents is also possible, because A. C. machinery can be insulated to stand the highest potentials. The A. C. dynamo has no commutator and the armature can in addition be made the stationary part. The A. C. induction motor (§538) also shares in the advantages of high insulation as well as efficiency and simplicity.

522. Two Phase and Three Phase A. C. Dynamos.—The alternating current dynamo is often made so as to generate more than one alternating current. The alternating currents in such machines always differ in phase, and so they are known as polyphase currents. The only systems in commercial use are the two and three phase systems. The currents are generated in coils placed at different angles on the armature. Thus in a two phase machine, there are two sets of armature coils, the first set cutting the magnetic field at a maximum rate, when the induced e.m.f. in the second set is zero, etc. Fig. 399 shows an arrangement of coils for a simple bi-polar machine by which two such currents could be generated. These currents differ in phase by a quarter period or 90°, and so a system using such currents is called a "quarter phase" system. Fig. 400 shows the phase of two such currents. Fig. 401 represents the phase relations of the three currents, where the phase difference is 120°. A machine wound to generate only one alternating current is called a single-phase machine, to distinguish it from the polyphase machines. The frequencies used with two and three phase alternating currents, are the same as with single phase alternating currents, that is, 60 alternations per second for ordinary conditions, with as low as 25 for power purposes alone. As already explained the commercial machines are always multipolar for mechanical reasons.

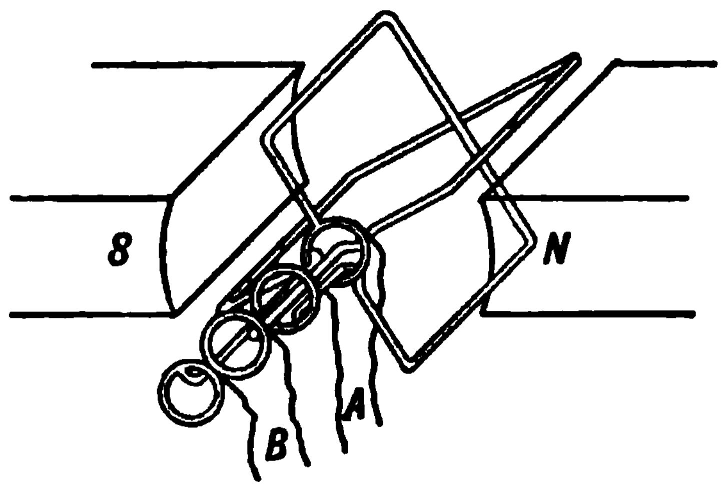


FIG. 399.

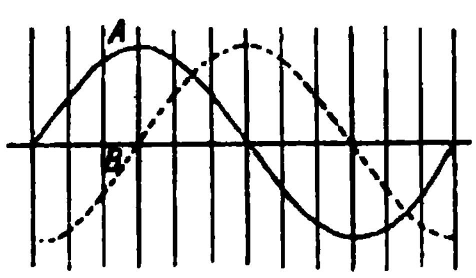


FIG. 400.

The advantage of two and three phase alternating currents is that the induction motor can be used. A polyphase machine has also generally a larger output for weight of machine than a single phase machine. The relative advantage of the two and three systems, lies in the size of conductors required for the dis-

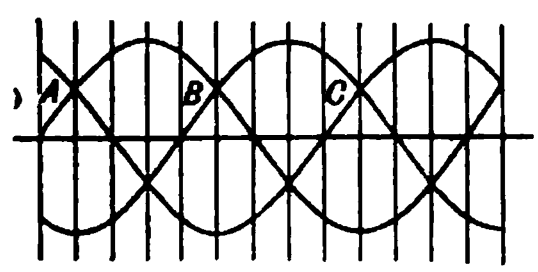


FIG. 401.

tribution of a given power. For a discussion of this, special treatises must be consulted.

523. Effect of Inductance in an A. C. Circuit.—When an alternating e.m.f. acts in a non-inductive circuit, the current I and the e.m.f. are in the same phase as is represented in Fig. 402. Here the current I at any instant is equal to E/R, where R is resistance of the circuit, and E is the e.m.f. in the circuit at the instant. When the e.m.f. is zero, the current is zero, and the maximum current corresponds in time to the maximum e.m.f., etc. If the circuit is inductive, then experiments with the oscillo-

Fig. 402.

Fig. 403.

graph (§524) show that the current lags behind the external or impressed e.m.f., that is, the current reaches a maximum later than the impressed e.m.f. This lag is due to self-induction. We have seen that starting or increasing a current in an inductive circuit produces an e.m.f. of self-induction which tends to retard the current growth, and similarly breaking or decreasing a current produces an e.m.f. of self-induction which tends to prolong the current. Thus the actual e.m.f. at any instant in an inductive circuit is the algebraic sum of the external or impressed e.m.f. (produced by generator, etc.), and the e.m.f. of self-induction. This e.m.f. of self-induction is equal to $-L(dI/dt)$ (§509).

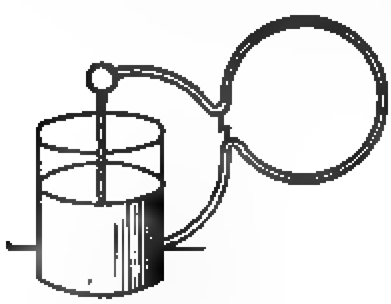


Fig. 404.

In Fig. 403 we have the curves of the impressed e.m.f. E, of the effective e.m.f., which is equal to IR, and of the e.m.f. of self-induction $-LdI/dt$, represented in their phase relations.

In the coil of an electromagnet, where the inductance is large, the e.m.f. of self-induction is correspondingly large, and this may be sufficient to make the effective e.m.f. practically zero. Such a coil is called a *choking coil.*

or an *impedance coil.* The resistance of an inductive circuit to an alternating current is called *impedance*. It can be shown that the impedance of a circuit is equal to $\sqrt{R^2+4\pi^2n^2L^2}$, where R is the resistance, L is the inductance of the circuit, and n is the frequency of the alternating current. The proof of this is given in treatises on alternating currents. This fact of impedance explains why little current goes through the primary of a transformer when the secondary circuit is not closed.

It is to be noticed that the impedance increases with the frequency n. The frequency of a Leyden jar discharge is ordinarily very high (§ 541), and so a discharge has a large impedance even through a single loop. Thus in Fig. 404 the discharge will leap across a considerable air-gap, sooner than go through the loop of wire.

524. The Oscillograph.—An ordinary galvanometer shows no deflection from an alternating current, because the needle system has so much inertia that it cannot follow the rapid impulse from the alternating current. Blondel, Duddell, and others have made galvanometers with very light moving parts and of high frequency, so that the needle system can follow the changes in the alternating currents. A galvanometer with a high frequency needle system, so that its deflections show the variations in alternating currents, is an oscillograph. Fig. 405 shows diagrammatically one of the best forms of oscillographs. It consists of a narrow loop of phosphor bronze strip, which is stretched with considerable tension, by a spring s, so as to have a very short natural period of vibration. This is in the strong magnetic field and placed with the plane of the loop parallel to the magnetic field. The strip is thus twisted by the magnetic forces when a current passes through the loop. The natural frequency of the loop is commonly from 3,000 to 10,000 vibrations per second, so that its deflections can follow very closely all ordinary variations in alternating currents. The deflections are recorded by a beam of light which is reflected from a small mirror M, attached to the loop. This beam of light falls on a photographic plate which

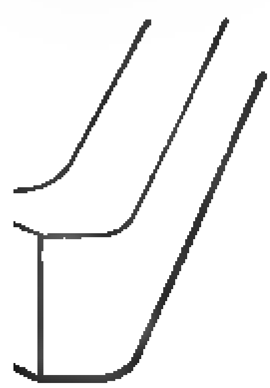


FIG. 405.

moves at right angles to the deflections. It thus leaves a curve showing the variations of the current.

Another form of oscillograph is the Braun cathode tube. The current passes through a helix, and thus acts magnetically on a pencil of cathode rays in a special form of Crookes tube (Fig. 406). When not deflected, this pencil of cathode rays produces a luminous spot on a phosphorescent screen in one end of the cathode tube. Under the action of an alternating current through the helix the luminous spot from the cathode rays vibrates in a luminous line on the phosphorescent screen. Looked at in a rotating mirror this is drawn out into an alternating current curve. The same may be photographed on a plate moving at right angles to the vibrations of the luminous spot.

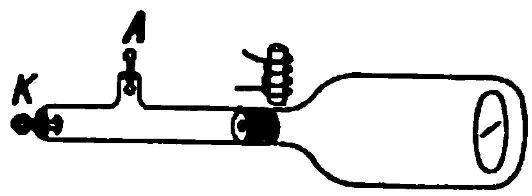


FIG. 406.

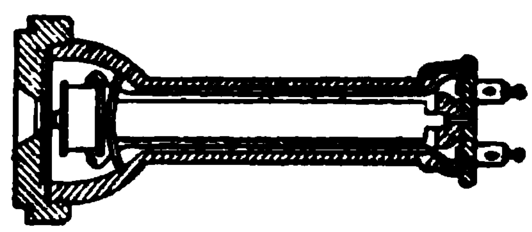


FIG. 407.

525. The Telephone.—The telephone, invented by Bell in 1876, consists of a thin iron plate or membrane, supported in front of the pole of a permanent magnet, and a spool of wire over the magnet pole (Fig. 407). Sounds can be transmitted electrically to a distance by using two telephones, one for a transmitter, and the other for the receiver. The two wire spools are connected in series by the wires joining the two stations. The sound waves set the thin iron plate in vibration, and the approach or receding of this plate changes the magnetic flux through the coil. This induces currents in the coils and the line which undulate in unison with and in proportion to the sound waves. These currents strengthen and weaken the attraction of the magnet of the receiver, and thus produce vibrations of the receiver plate which correspond to the vibrations of the transmitter plate. The electric currents induced in the above cases are very feeble, and can transmit sounds only short distances. For longer distances, the microphone described in the next section is used as a transmitter.

526. The Microphone.—The microphone depends upon a fact discovered by Hughes in 1878, that the electrical resistance of a loose contact between two conductors changes under the action

of sound waves. Variations of the current can thus be produced in a circuit, these variations corresponding to the sound waves which produce them. A simple form of microphone consists of a piece of carbon resting on two pieces of carbon, and thus completing a circuit which includes a battery and a Bell receiver.

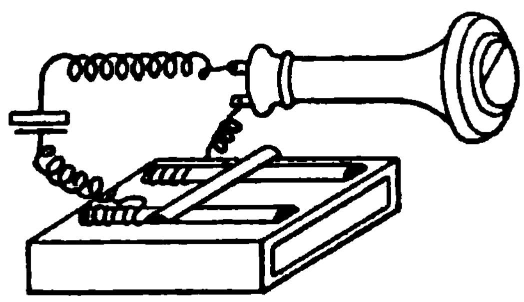


FIG. 408.

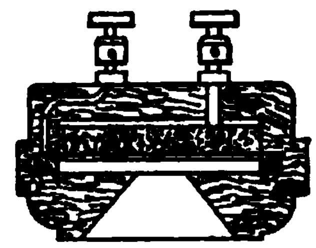


FIG. 409.

The carbons can be mounted on a sounding box. Such an arrangement makes an effective transmitter (Fig. 408). In the Hunning's transmitter, which has been extensively used in long distance telephony, granulated carbon is placed in between two metal plates as shown in Fig. 409.

ELECTRODYNAMICS

527. Motion of a Circular Circuit in a Magnetic Field. *Maxwell's Rule.*—If a conductor carrying a current is placed in a magnetic field (§ 368), there is in general a force tending to move the conductor. It has been shown (§427) that a circular circuit or other plane closed circuit, acts like a magnet, and that the lines of force due to the current enter the face in which the current flows clockwise, and emerge from the face in which the current flows anti-clockwise (Fig. 410). When such a coil is placed between the poles of a magnet, the coil tends to place itself so that its plane is at right angles to the field, the clockwise face of the coil being toward the *N* pole; in other words, the coil places itself so that the lines of force of the field and of the coil coincide. Maxwell has generalized this into a rule—*An electric circuit tends to*

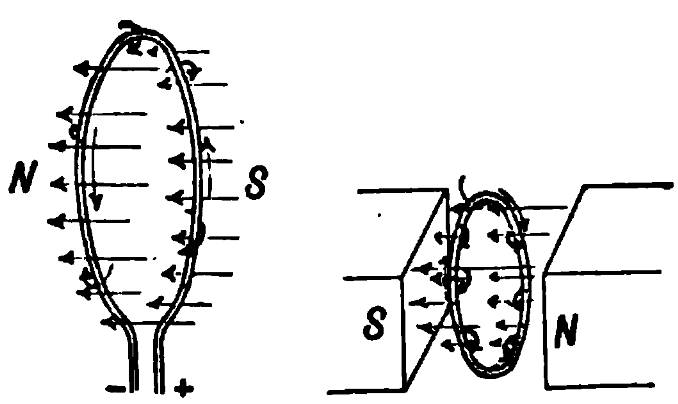


FIG. 410.

move in a magnetic field so as to include the maximum number of lines of force. Thus the lines of the field and of the circuit are in the same direction when stable equilibrium is reached

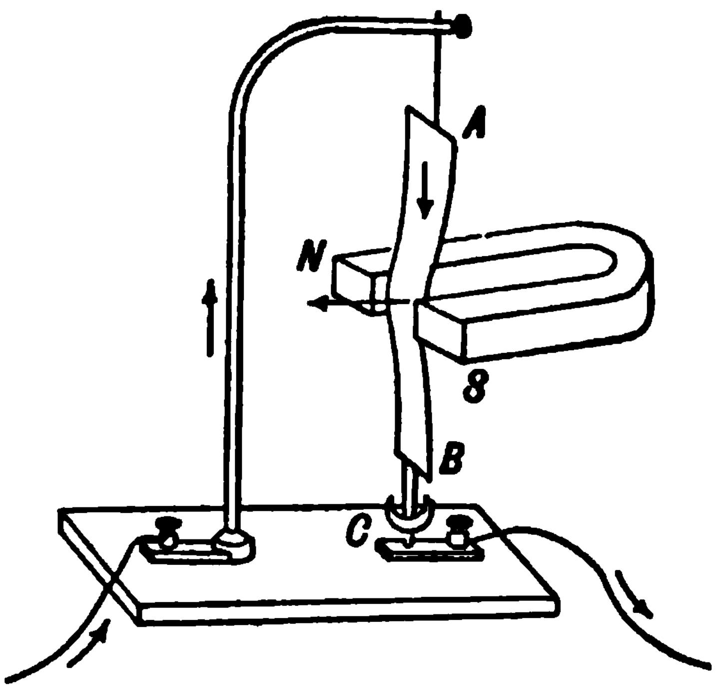


Fig. 411.

528. Force on a Linear Circuit in a Magnetic Field.—In Fig. 411 is represented an experiment for showing the action between a linear circuit and a magnetic field at right angles to the circuit. *AB* is a strip of flexible copper foil hanging vertically and making connection with the mercury cup *C*. The current flows from *A* to *B* as indicated. The magnetic field is that of a U-shaped magnet and is horizontal. The flexible conductor is acted on by a force at right angles to the plane of the field and of the current. The relative directions of the current, the field and the motion, are shown in Fig. 412. Fleming has given the following convenient rule for remembering these directions, the rule being similar to that for induced currents (§502). Hold the *left* hand with the thumb, the forefinger and the middle or center finger, so that each is at right angles to each of the others; then, if the forefinger is in the direction of the field and the center finger in the direction of the current, the thumb will indicate the direction of the resulting force on the linear circuit.

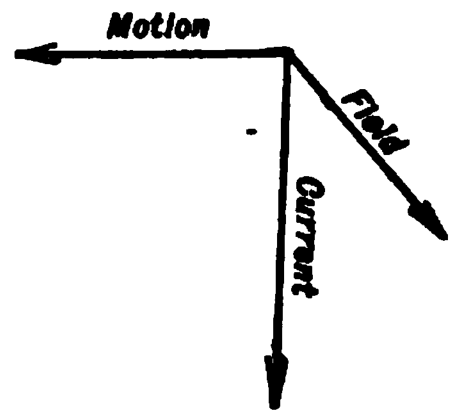


Fig. 412.

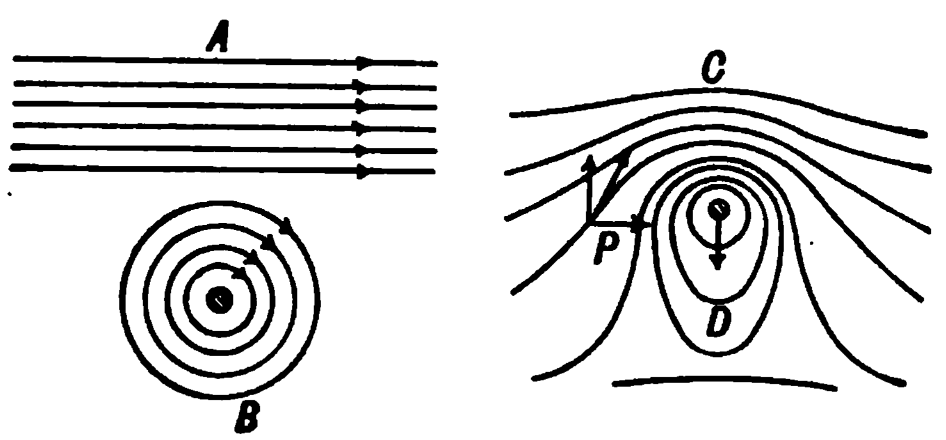


Fig. 413.

Fig. 413 shows the distribution of the magnetic lines about a current which is at right angles to a uniform magnetic field (A). The current flows down, at right angles to the plane of the figure, and hence its magnetic lines are clockwise (B). This is compounded with the uniform

magnetic field, strengthening the field on the side C and weakening it on the side D. The movement of the circuit in the direction C to D, can be considered as due to the tendency of the lines on the side C to contract. The addition of the fields is indicated at P.

529. Numerical Value of the Force on a Linear Circuit in a Magnetic Field.—We have seen in §428 that a circuit element ids acts on a magnetic pole m at a perpendicular distance r with a force $F=\frac{mids}{r^2}$. There is evidently an equal and oppositely directed force acting on ids due to the field on m, that is a force $-F$, which is at right angles to the plane of the pole and the circuit. The field due to m is $H=m/r^2$. Hence the force on ids is $F=Hids$. For a circuit of length L, $F=HiL$. That is, *a linear circuit Li at right angles to a uniform magnetic field H is acted on by a force of HiL dynes*, and this force is at right angles to the plane of H and iL. If iL is not at right angles to H, $iL \sin\theta$ the component of iL at right angles to H is to be taken, where θ is the angle between H and iL.

FIG. 414.

530. Force Between Two Parallel Circuits.—If two circular coils carrying currents are hung by flexible wires parallel to each other (Fig. 414) they attract each other when the currents in the two coils are in the same direction, and repel each other, when the currents are in opposite directions. It is easy to see that each of the circuits thus tends to move so as to include the maximum number of lines of force. Rectangular or other plane closed circuits can be substituted for the circular circuits.

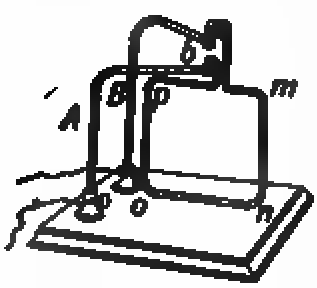


FIG. 415.

Lord Kelvin makes use of the forces between parallel circuits in his electrodynamic balance for measuring electric currents.

531. Force Between Two Circular Currents at Right Angles.—Two circular currents at right angles to each other, tend to rotate and place themselves in the same plane, and with the currents in the same direction. Ampère, who was the first to study the actions of currents on currents, used his "electrodynamic apparatus" shown in Fig. 415. The two mercury cups a and b, are at the ends of the conducting

supports A and B, and a is vertically over b. The wire frame *mnop* is bent so that its ends dip in the mercury cups, and the wire frame is thus free to rotate about a vertical axis through ab.

The force of rotation between coils is used in the electrodynamometer for measuring currents. Fig. 416 shows the Siemens' form of the electrodynamometer. The coils are kept at right angles to each other by putting more or less torsion in a spiral spring which is attached to the movable coil. The amount of torsion is read by a torsion head and circle on top. The force between the coils is proportional to the square of the current, and the opposing force of torsion is proportional to θ, the angle of torsion. Hence $I^2 = k^2\theta$, or $I = k\sqrt{\theta}$, where k is a constant to be determined for each instrument. This instrument is adapted to the measuring of both direct and alternating currents.

FIG. 416.

632. Force between Parallel Linear Circuits.—By using his electrodynamic apparatus, Ampère was able to show that two parallel straight circuits attract each other when the currents are in the same direction. An arrangement of the apparatus to demonstrate this, is shown in Fig. 417. In Fig. 418 A and B are normal sections across the parallel conductors, and the circular lines of force are shown for the case of both currents flowing away from the reader. Evidently the maximum number of lines of force for each circuit exists when A and B are close together. Hence we should expect attraction from Maxwell's rule. The repulsion between oppositely directed currents follows similarly.

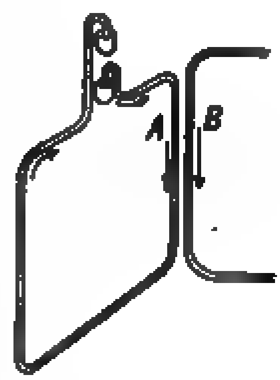


FIG. 417.

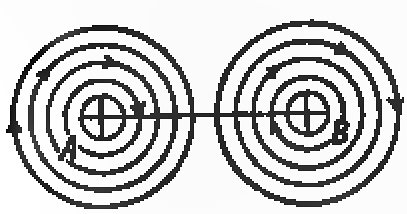


FIG. 418.

Ampère extended the above law as follows: Two straight conductors which cross each other obliquely, attract each other when their currents both flow toward or both away from the point of crossing; when one current flows away and the other toward the crossing point, they repel each other.

533. Roget's Spiral.—An interesting case of the attraction between parallel circuits is that of Roget's spiral (Fig. 419). A spiral brass coil *AB* hangs vertically, and its end *B* dips in a mercury cup *M*. When a current passes through *AB* the parallel coils attract each other and lift *B*, thus breaking the circuit; the attraction ceases and contact at *B* is again made. This repeats itself indefinitely.

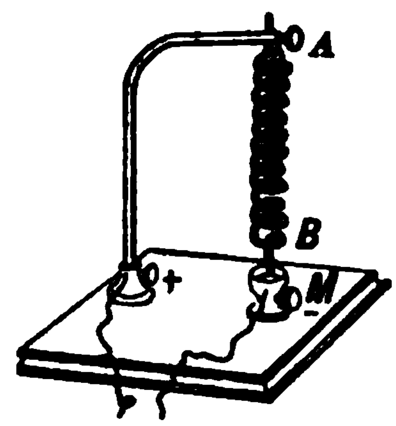


FIG. 419.

534. Electromagnetic Rotation.—The first apparatus to produce rotation of a conductor from electric currents and magnets was described by Faraday in 1821. A glass tube, Fig. 420 is stopped at both ends with corks, and the lower cork is covered with mercury, and has one pole of a magnet *NS* stuck through it. A platinum wire hangs by a hook from the upper cork and dips in the mercury below. When a current flows through the *platinum* wire the wire revolves about the magnetic pole, and continues as long as the current flows. Reversing the current reverses the direction of the rotation. This is evidently a case of a conductor moving at right angles to the plane of the current and of the magnetic field from the pole (§528). Fig. 421, shows a common form of an electromagnetic rotation apparatus. Its action can be followed from the above.

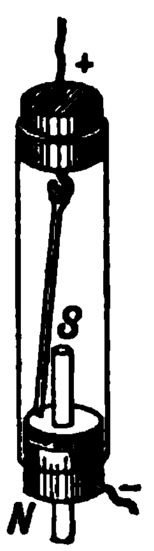


FIG. 420.

535. Barlow's Wheel.—This consists of a metal disk mounted to revolve about a horizontal axis over a mercury trough (Fig. 422). The lower edge of the disk touches the surface of the mercury.

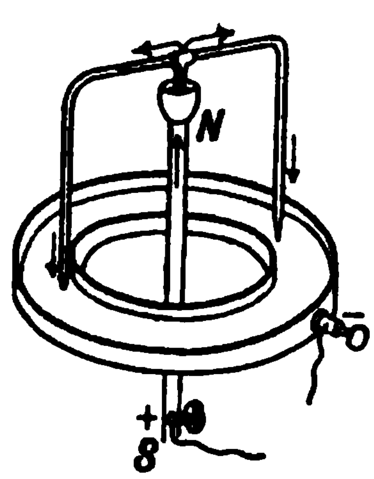


FIG. 421.

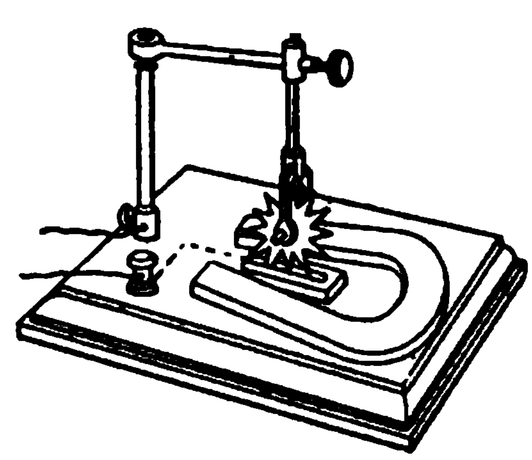


FIG. 422.

cury. A U-shaped magnet lies so that its lines of force cut the lower half of the disk at right angles. When an electric current flows from the axis through the disk to the mercury, the disk

rotates. Reversing the direction of either the current or of the magnetic field reverses the direction of rotation. A star-pointed wheel as shown in the figure is often used instead of the solid disk, in order to reduce the friction at the mercury surface. Each radius of the wheel in turn as it dips in the mercury becomes a part of the circuit, and is acted upon by a force at right angles to plane of the magnetic field and the current (§528), and thus causes continuous rotation. Barlow's wheel and Faraday's disk (§506), are evidently inverse machines, the first using electrical energy to produce mechanical motion, and the second using mechanical energy to produce electrical energy.

536. Direct Current Motors.—The direct current dynamo becomes a motor when an external current is sent through its field magnets and armature. It then transforms electrical energy into the mechanical energy of the rotation of the armature. The forces acting on the armature circuits follow laws already treated in the sections on the motion of circuits in a magnetic field.

The rotation of the armature in the magnetic field induces in the armature coils an *e.m.f.* which opposes the current driving the machine. This back *e.m.f.* increases with the speed of the motor. Thus the current through the motor decreases as the speed increases. When the speed causes a back *e.m.f.* equal to the impressed *e.m.f.*, there is no current through the armature. This occurs when a frictionless motor runs under no load, and is accordingly doing no work. At starting, there is no motion and no back *e.m.f.*, and hence the current is a maximum. To prevent injury from the "rush" of the current before the motor reaches a speed to produce a back *e.m.f.*, a starting resistance is commonly placed in series with the armature. This starting resistance is gradually reduced as the speed of the motor increases.

537. Alternating Current Motors.—Two similar single-phase A. C. machines can be used as generator and motor, provided the motor is first brought to a synchronizing speed with the generator. An A. C. machine thus used as a motor is called a synchronous motor. The synchronous motor is efficient, but has the great disadvantage of not being self-starting, and also of stopping if the motor is thrown "out of step" by overloading. Hence synchronous motors are not in general use. The successful use of alternating currents for power transmission has been due to the

nvention of the polyphase induction motor. The principle of this motor was first discovered and stated by Ferraris, in 1885, and its application was developed by Tesla and others. Both two and three phase currents have been successfully used in these motors.

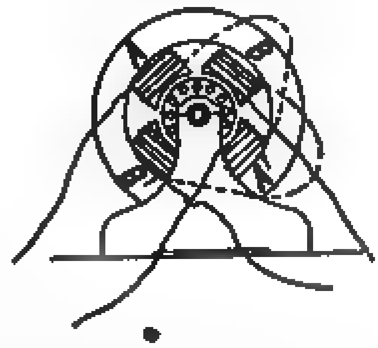


Fig. 423.

638. Simple Two Phase Induction Motor.—In Fig. 423 we have represented two sets of helices AA' and BB', placed at right angles to each other. An alternating current through the helix AA' produces an alternating magnetic field in the line AA'. If the current through AA' is sinusoidal, represented by the equation $i_A=I\sin pt$ (§ 514) then the intensity of the field is represented by $h_A=NI\sin pt=H\sin pt$, where N is a constant depending upon the number of windings on AA', etc. Similarly, an A. C. in BB' will produce an alternating magnetic field in the line BB'. Suppose the A. C. in BB' to differ in phase from that in AA' by a quarter phase, that is, $i_B=I\sin(pt+90°)=I\cos pt$. Then the field BB' is $h_B=NI\cos pt=H\cos pt$.

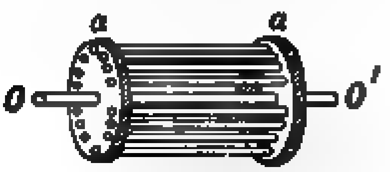


Fig. 424.

The field at any instant is thus the resultant of the two component fields, $h_A=H\sin pt$, and $h_B=H\cos pt$. The resultant

$$F=\sqrt{h_A^2+h_B^2}=\sqrt{H^2(\sin^2 pt+\cos^2 pt)}=H$$

That is the resultant is constant in strength. It evidently rotates with every complete alternation of the current.

Suppose we place in this rotating magnetic field, a "squirrel cage" armature, which can rotate about its axis OO', which is perpendicular to the field.

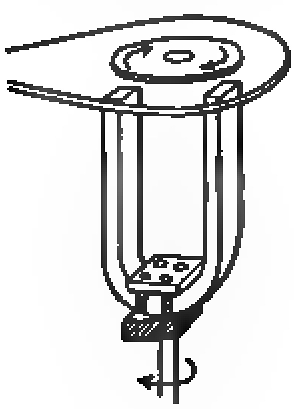


Fig. 425.

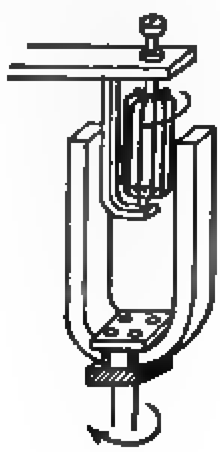


Fig. 426.

Fig. 427.

The construction of this "squirrel cage" armature can be seen from Fig. 424. It consists of two copper disks mounted on a shaft OO', and with copper bars around the circumference of the disks, joining the disks. It is found

that such an armature rotates approximately synchronously with the rotating magnetic field. The explanation is simple. If the armature were held still, there would be electric currents induced in the bars. By Lenz's law these induced currents would react magnetically on the rotating field, with a force tending to prevent the relative motion of the armature and the field. In other words, there is a torque causing the armature to rotate with the field. Figs. 425–7 show a series of simple experiments for demonstrating the principle of the rotating field, and the induction motor. In Fig. 425 we have an aluminum disk, mounted on a pivot. Under this is a magnet mounted on a whirling table so that it can be rotated. The disk rotates with the magnet. This is the inverse of Arago's disk (§507), and the explanation by electromagnetic induction is the same. In Fig. 426, the experiment is varied by substituting a "squirrel cage" cylinder, pivoted on a vertical axis between the poles of the rotating magnet *NS*. In Fig. 427, we substitute for the rotating magnet, two coils, placed with the planes at right angles. By passing through the coils alternating currents, which differ in phase by a quarter period, a rotating magnetic field is produced, and the "squirrel cage" rotates.

ELECTRIC OSCILLATIONS AND WAVES

539. Electric Oscillations.—In the preceding sections, we have described alternating electric currents with frequencies which are commonly between 25 and 125 per second. It has been seen that these alternating currents follow special laws which are due to the special importance of inductance in such circuits. At still higher frequencies, alternating currents bring in new phenomena with additional laws. Alternating currents of high frequency are called oscillatory currents, or **electric oscillations.** The lowest frequency for which the term oscillatory is used is naturally not definite, but we may in general think of an electric oscillation as having at least 1,000 alternations per second. It is often several millions per second.

The study of electric oscillations has been in recent years one of the most important and fruitful in physics. It has led to the discovery of electromagnetic disturbances in the space about oscillatory currents, disturbances which are propagated outward as electric waves. These electric waves have been shown to be identical physically with light waves, except in being of longer wave-length. Heinrich Hertz, the discoverer of electric waves, was thus able to prove experimentally the theory of James

Clerk Maxwell, that light is an electromagnetic phenomenon (§543). The experiments of Marconi and others have resulted in using electric waves to transmit signals by the electric-wave telegraphy. Lodge and his fellow workers have also explained many of the "mysteries" of lightning discharges by laws proved for oscillatory currents.

540. Methods of Generating Electric Oscillations, Alternators.—Two general methods of producing high frequency electric currents or oscillations have been used, (*a*) by multipolar alternating dynamos, and (*b*) by an electric discharge in a circuit containing capacity and inductance in certain ratios, with low ohmic resistance.

A high frequency dynamo-electric machine must have a large number of poles, and be driven at a high velocity. Such machines have been constructed by Tesla, Ewing, Duddell and others. Frequencies of from 10,000 to 15,000 per second were ordinarily reached, and in one machine a frequency of 120,000 per second is recorded. But the velocity of the moving parts must be so high that only small machines are mechanically possible. The high frequency alternator has only recently been developed as a generator of electric oscillations, and then only for special work.

The generally used method of producing oscillations of the highest frequency is based on the oscillatory character of the discharge of a circuit containing capacity. This will be described in the next section.

541. Oscillations by a Condenser Discharge.—When a Leyden jar is discharged, there is a flash which to the eye appears as a single spark. But as early as 1842 Joseph Henry concluded that this discharge of a Leyden jar "is not correctly represented by the single transfer of an imponderable fluid from one side of the jar to the other." "The phenomena," he continues, "require us to admit the existence of a principal discharge in one direction, and then several reflex actions backward and forward, each more feeble than the preceding until equilibrium is obtained." Henry reached this striking conclusion by observing the irregular magnetization of steel needles by Leyden jar discharges. Henry's conclusion was confirmed by the mathematical theory of Lord Kelvin, published in 1853. Kelvin showed that the

character of the discharge depended upon the resistance R, the capacity C, and the inductance L, and that the frequency is given by the equation,

$$n=\frac{1}{2\pi}\sqrt{\frac{1}{LC}-\frac{R^2}{4L^2}}$$

If $R^2/4L^2$ is so small as to be neglected compared to $1/LC$, then the frequency is

$$n=\frac{1}{2\pi\sqrt{LC}}$$

or the period $T=2\pi\sqrt{LC}$. That is, if the resistance of the discharging circuit is small, then the discharge is oscillatory. These oscillations are rapidly damped. When the resistance R is large then the term under the radical is negative and the frequency becomes imaginary. The discharge is unidirectional, dying away slowly. Figs. 428, a and b, are curves showing these two types of discharge.

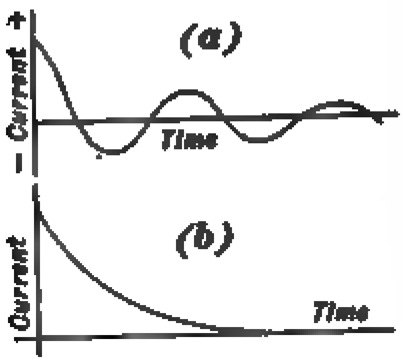


Fig. 428.

In 1858 Feddersen confirmed Kelvin's theory, showing by examining the spark discharge with a revolving mirror that the spark consists of a series of alternating and diminishing flashes. Others have photographed these flashes. One of the most beautiful confirmations of the oscillatory character of Leyden jar discharge is shown in the photograph reproduced in Fig. 429. This was made by Zenneck in 1904, using a Braun tube as an oscillograph (§524.)

The discharge of a condenser is thus analogous to the vibrations of an elastic rod clamped at one end. When bent and released, the rod in general vibrates back and forth about an equilibrium position, dissipating its energy, and finally coming to rest. But if the rod is immersed in a heavy oil, which offers considerable resistance to motion, the rod comes slowly to rest without vibrating beyond its equilibrium position.

Fig. 429.

542. Electric Oscillations and Waves. Resonance.—In 1888 Heinrich Hertz showed that a conducting system in which electric oscillations are produced becomes the source of electric waves, and that these waves can be detected by oscillations set up in a similar circuit called a resonator. Fig. 430 shows one of Hertz's arrangements. The discharge rods *A* and *B* are connected to terminals of the secondary of the induction coil *C*, and are separated by the discharge gap *P*. The metal spheres *S* and *S′* slide on the rods, so that the length of the discharge circuit can be varied. For a receiving circuit or resonator, Hertz used a loop of wire *R* broken by the spark gap *P′*. He found that when the

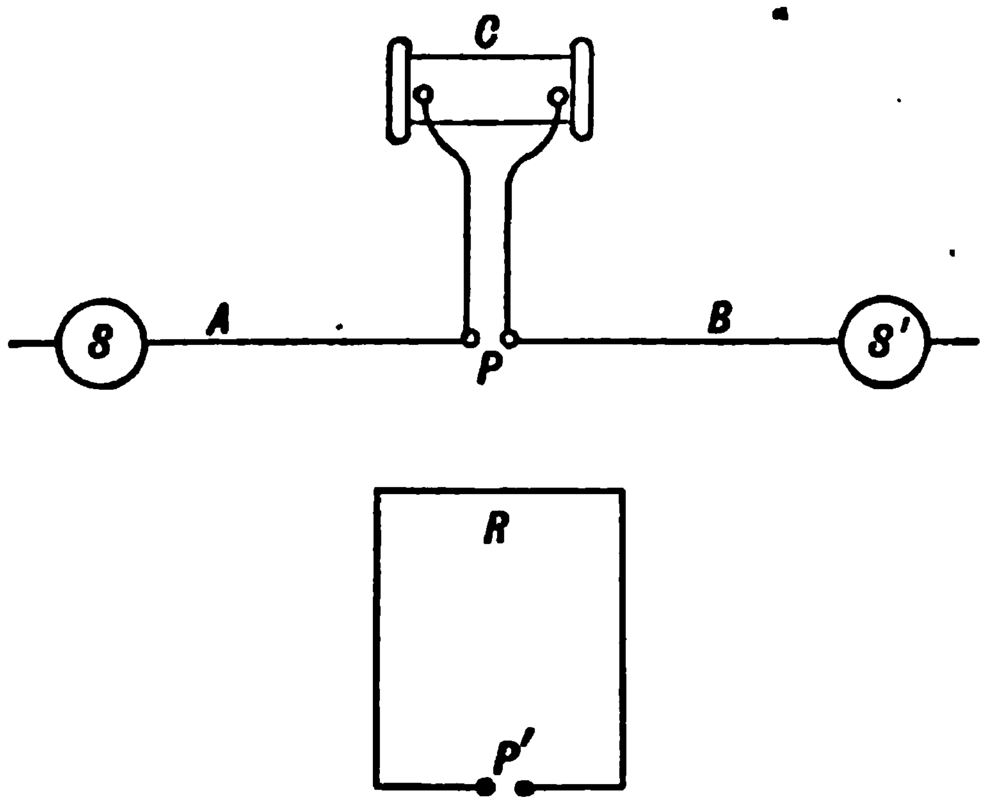


FIG. 430.

two circuits were "in tune," a discharge at *P* caused a spark at *P′*; or in other words, oscillations in the first circuit produced oscillations in the second circuit. The explanation is evidently exactly like that of the experiment of resonance between two tuning forks on resonators. The sound waves sent out from a tuning fork *A* set in vibration a second fork *B*, provided the two forks are of the same pitch. The electric waves from the oscillator produce the electric oscillations in the resonator, provided they are "in tune." Fig. 431 shows a striking class-room experiment due to Lodge for showing electrical resonance. *A* and *B* are two equal Leyden jars. The jar *A* has a wire loop *L* which forms the discharge circuit, the gap being between the polished balls at *P*. The jar is charged by a small static electric machine.

The inner and outer coatings of the jar *B* are connected by a wire loop *L'*, the inductance of which can be varied by the sliding wire *M*. By using a tin-foil strip, a small gap *G* is left between the inner and outer coatings of *B*. When the two circuits are in tune, a discharge in *A* produces oscillations in *B*, which are shown by a bright spark at *G*.

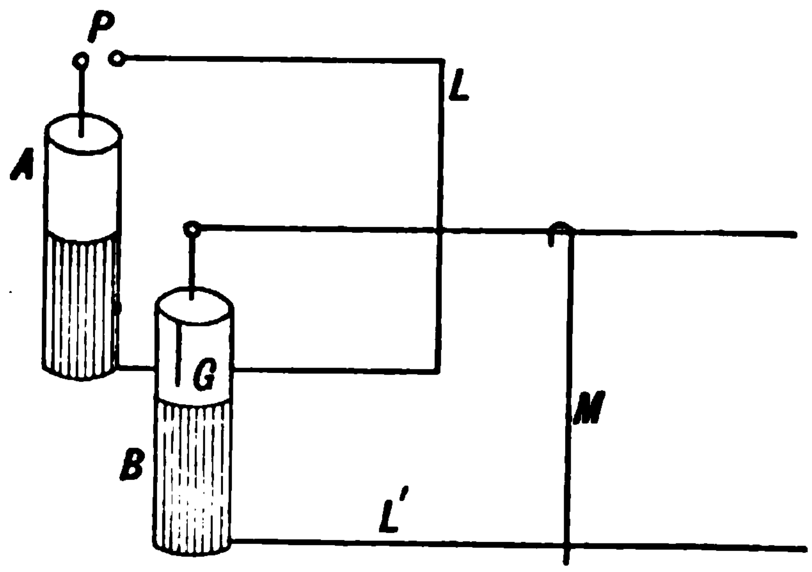


FIG. 431.

543. Electromagnetic Theory of Light.—Using his spark gap detector, Hertz showed that electric waves are reflected from plane and curved metal surfaces in accordance with the same laws as light waves; that they are refracted in passing through prisms of resin, paraffine and other dielectrics; that they are polarized by a coarse metal grating, and hence are transverse waves. He measured their wave-length and computed from his oscillator their frequency; and thus, from the formula $v = n\lambda$, he determined that the velocity of electric waves is the same as that of light. The electric waves, which Hertz produced, generally had wave-lengths of eight or nine meters. The shortest electric wave yet produced has a wave-length of about four millimeters, still many times the length of the longest infra-red line (§733).

Twenty years before Hertz's experiments were performed, Maxwell advanced the view that waves of light are electromagnetic waves of very short wave-length. From theoretical calculations Maxwell found that the velocity of such waves equal $1/\sqrt{k\mu}$, where k is the dielectric constant of the medium and μ its permeability, both being expressed in electromagnetic units. The velocity thus calculated for air agrees with the velocity of light (§644). The value of μ for transparent substances is nearly 1.

Hence the index of refraction (§667) from a substance of dielectric constant k_1 to another of dielectric constant k_2 is $n = \sqrt{k_2/k_1}$. This relation has also been verified in many cases, but the dependence of n on the wave-length makes the test difficult in other cases. The waves started from Hertz's oscillator (§542) are *plane polarized.* At P' there is an alternating electrostatic force in the plane of the diagram and an alternating magnetic force per-

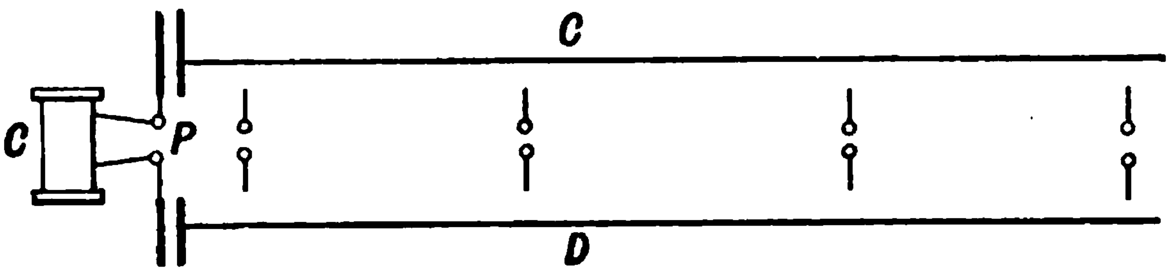


FIG. 432.

pendicular to that plane. These together constitute the vibration in the front of the wave and a plane polarized wave of light is similarly constituted. Thus the electromagnetic theory supplements the wave theory stated in Light, by explaining the nature of the wave-motion.

544. Electric Waves along Wires.—Fig. 432 shows a form of Hertz oscillator as modified by Lecher to show electric waves along wires. The oscillations produced by the discharge across

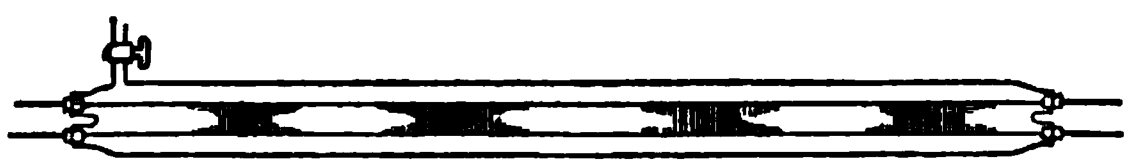


FIG. 433.

P, act by static induction, and produce waves which traverse the wires C and D and are reflected back, thus forming standing waves by interference between the advancing and the reflected waves (§253), similar to the standing waves in organ pipes (§609). The nodes and loops can be detected by sliding a small gap along the wires, or easier by a device due to Arons, shown in Fig. 433. Arons enclosed the two wires in an exhausted glass tube. The loops are indicated by the electrical discharges, while the nodes remain dark.

Seibt has arranged a beautiful class-room experiment (Fig. 434) in which he uses a Tesla transformer T (§512) as oscillator, and

a special resonance coil *CD* to show standing waves. The vertical coil *CD* is about two meters high and consists of a coil of silk-covered wire on a wooden core. Parallel to it and insulated from it, is a stretched wire *MN*. The nodes and loops come out brilliantly in a darkened room as indicated in Fig. 434*b*.

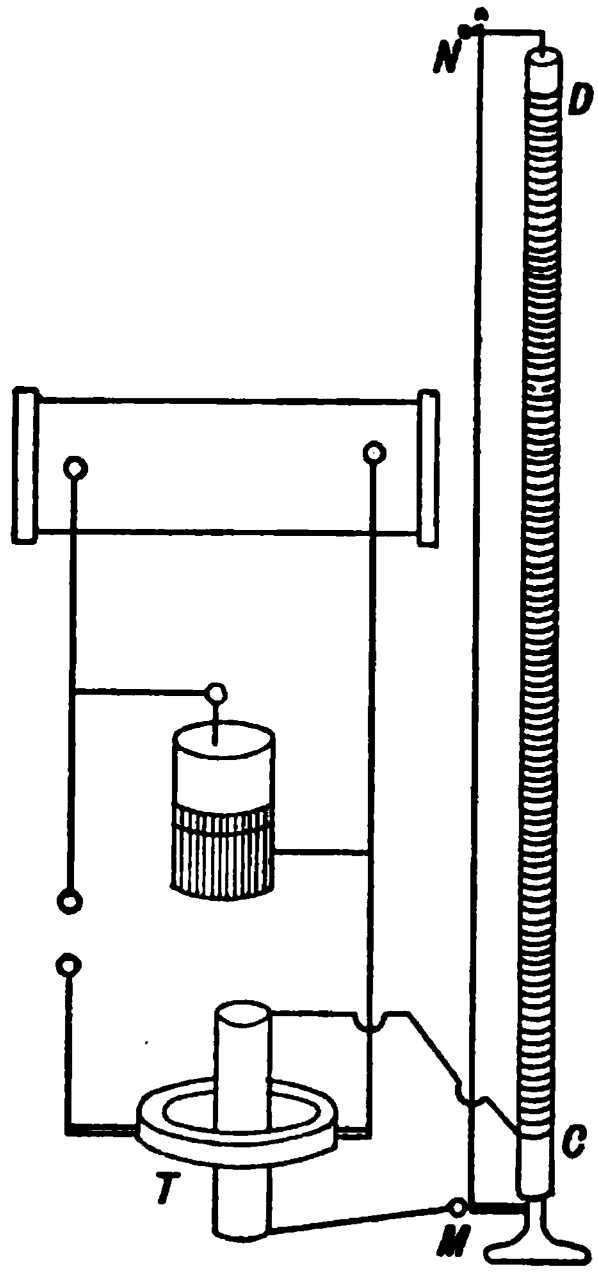


Fig. 434*a*.

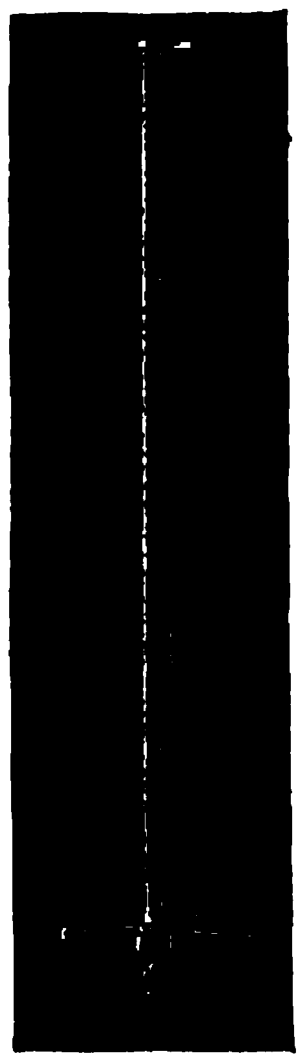


Fig. 434*b*.

545. Detectors of Electric Waves.—The spark gap, which Hertz used so successfully in his investigations, has been largely replaced by more sensitive detectors. (Cymoscope has been proposed as a general name for electric wave detectors.) Almost every effect of an electric current has been used in these detectors, such as heating, magnetic, electrolytic and resistance effects. Only two of these detectors, the coherer, and the crystal-rectifier will be described here. The reader is referred to special treatises for accounts of the others.

The coherer, in the form given to it by Marconi, consists of a small glass tube *TT'* Fig. 435, in which there are two silver electrodes *PP'*, separated by a small quantity of loosely packed

metal filings. A mixture of 95 per cent. nickel and 5 per cent. silver filings has been successfully used by Marconi. Marconi also found that exhausting the tube of air increased the reliability of the coherer. The action of the coherer depends upon a discovery made by Branly in 1900. He discovered that loosely packed metal filings, which offered practically infinite resistance to an electric current, suddenly acquire good conductivity under the action of an electric wave. When lightly tapped or shaken, the filings again lose their conductivity. The generally accepted explanation is that the small filings cohere owing to the welding action of the infinitesimal sparks produced by the electric wave, and hence the name coherer was given. The coherer is not selective in its action, that is, it responds to electric waves of many or all lengths. The method of using the coherer can be seen from the diagram in the next section.

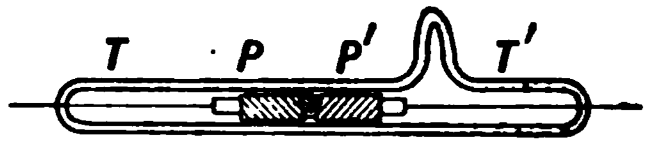


FIG. 435.

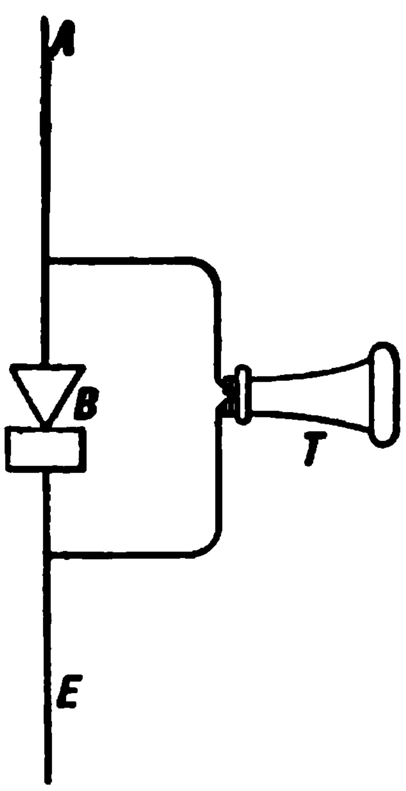


FIG. 436.

It has been found that certain crystals, such as silicon, molybdenite, and carborundum, have the property of offering much greater resistance to the passage of one-half of a rapidly alternating electric current than to the other half. If such a crystal is included in the circuit of a receiving antenna, it allows the electric oscillations in one direction to pass, but practically cuts out the oscillations in the opposite direction; that is, it "rectifies" a train of rapidly alternating oscillations into a train of unidirectional pulsations. These trains of electric pulsations can be detected by the clicks in a telephone circuit which is connected with the antenna. In Fig. 436 is shown the connections; *B* represents the crystal rectifier, *ABE* part of the antenna circui and *T* a telephone.

546. Electric Wave Telegraphy and Telephony.—Since 1895, Marconi has developed a system of electric wave telegraphy, more often called wireless telegraphy, for transmitting signals to a distance. Using very powerful oscillators and extremely sensitive detectors, Marconi has transmitted messages thousands of

miles. This system has been particularly successful in communicating with and between ships at sea. Fig. 437 shows a diagram of a very simple electric wave telegraphic arrangement. A and A' are high vertical lines, called antennæ. P is the spark gap of the sending station, C is the coherer, R is a relay operated by any current through C. This throws in the battery B_2, and excites the magnet M which decoheres C by tapping it. E and

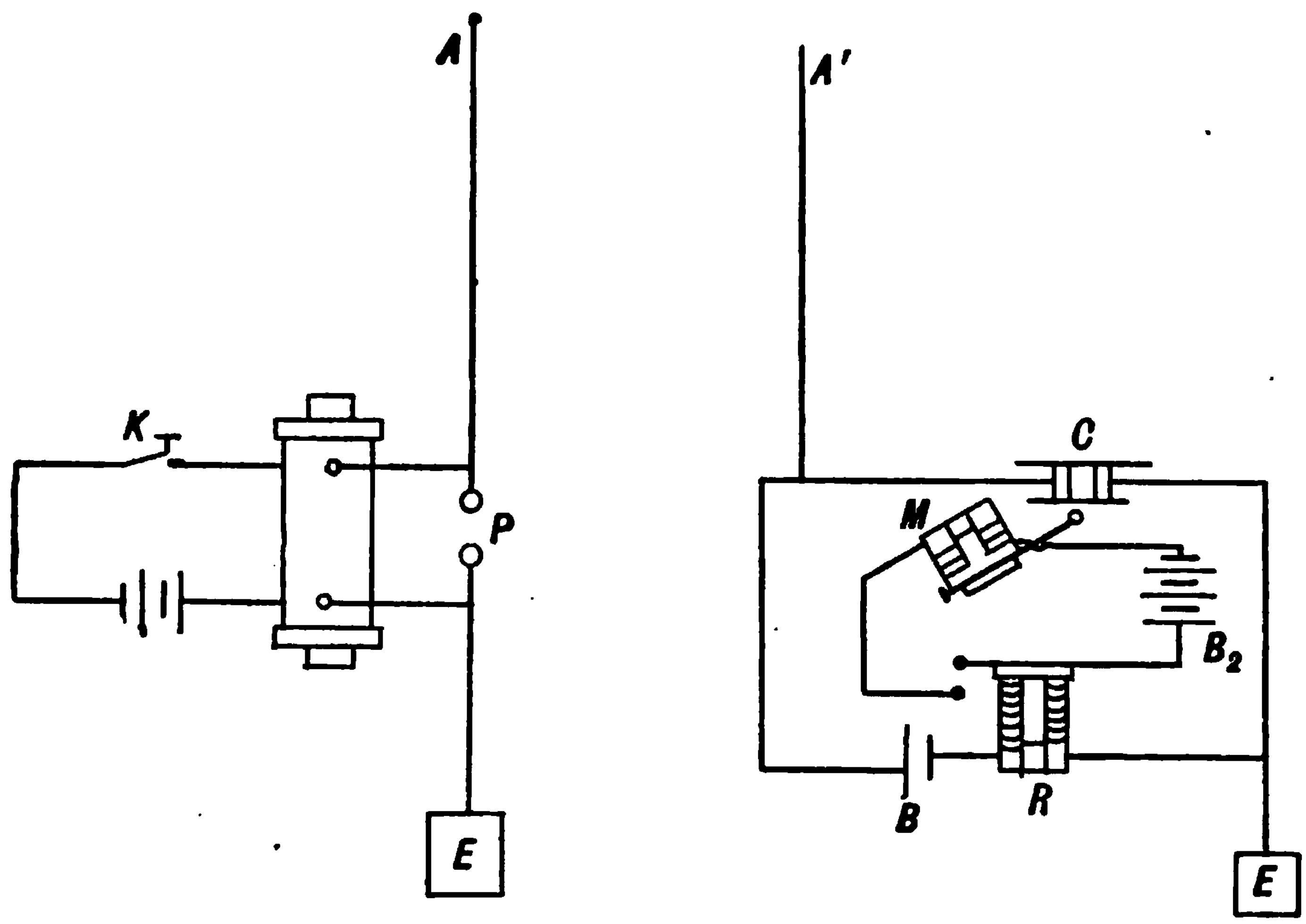


FIG. 437.

E are earth connections. Instead of the coherer C and sounder, more reliable and more sensitive detectors are now used.

One of the most important developments in electric wave telegraphy has been in the methods of producing trains of undamped waves. The difference between damped and undamped waves is indicated by Fig. 438*a*.

The waves produced by an ordinary spark discharge are as we have seen damped waves (see Fig. 429). The simplest means of producing undamped waves is by an alternating current dynamo; but the design of such a machine, so as to have a high frequency with sufficient output of electrical energy, is not easy because there is a limit to the speed at which an armature can be safely rotated. An entirely different method of producing undamped electric waves is that first patented by

Elihu Thomson in 1892. This is shown in Fig. 438*b*. The circuit of the electric generator *D* consists of a coil *I* of high inductance, of a condenser *C*, an inductance *L*, and in parallel

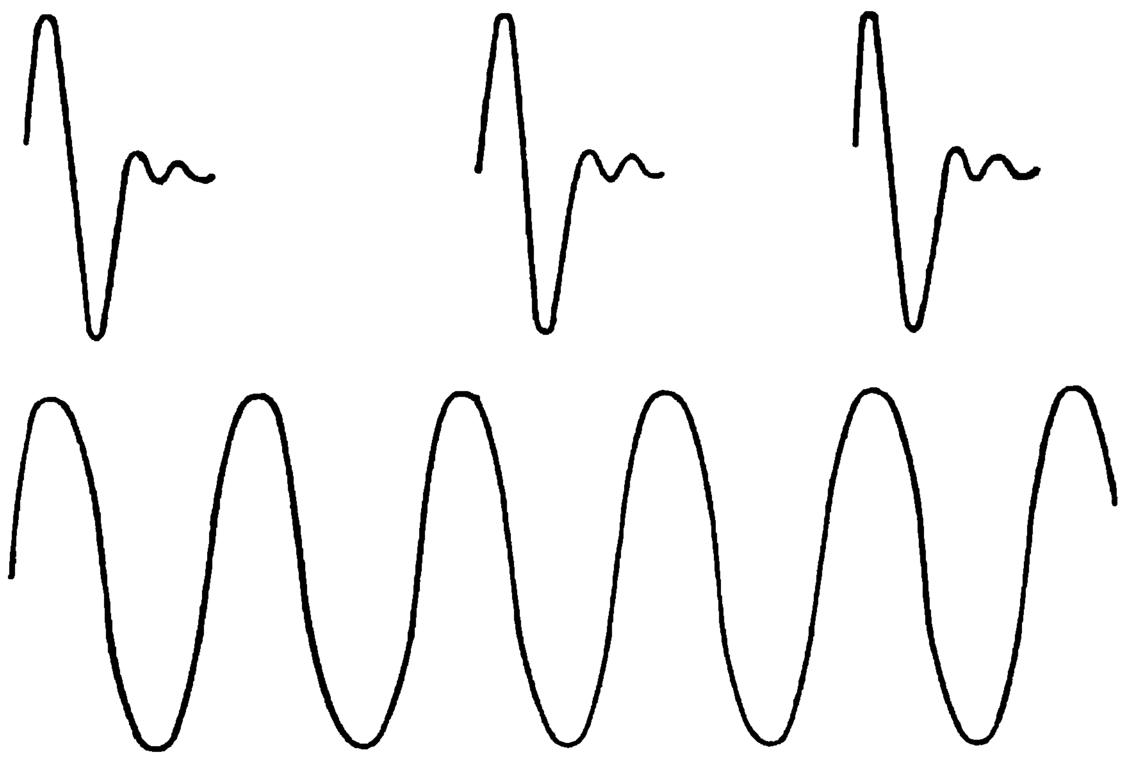


FIG. 438*a*.

a spark gap *S*. When a spark discharge passes across *S*, persistent electric oscillations are set up in the circuit, and hence undamped waves are sent out. Duddell substituted hard carbons for the balls S of the spark gap, and thus produced the "singing arc." Poulsen has developed the method still further by using a hydrocarbon gas about the arc, making the positive electrodes of copper, etc. Several other systems of undamped waves are also in use. A great advantage of undamped waves in wireless transmission is that more power can be used and longer distances can be covered. The most striking advantage is, however, that it makes wireless telephony possible, if the frequency of the undamped waves is 20,000 or over. This frequency produces a note so high that it is not audible. The lower frequencies of the voice are now superimposed, by introducing a telephone transmitter, as variations of the undamped electric waves, and these variations are reproduced in the distant receiving apparatus, without interference from the fundamental waves. In 1915, a notable

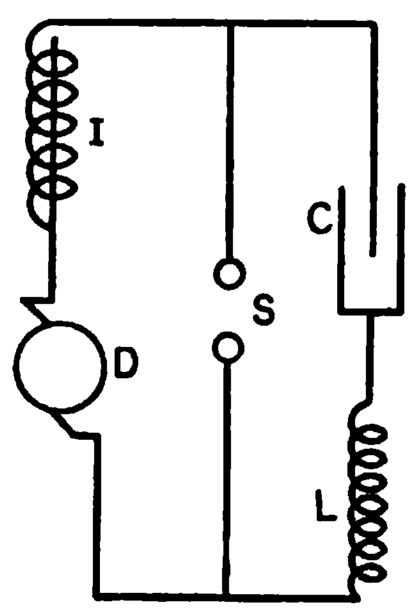


FIG. 438*b*.

advance was made in electric wave telephony by transmission of speech between Washington, D.C. and Eiffel Tower in Paris.

DIMENSIONS OF ELECTRICAL UNITS

547. Kinds of Electrical Units.—Three kinds of eletrical units have been defined and used in the previous sections, the electrostatic units, the electromagnetic units, and the "practical" units. The practical units have been defined as multiples of the electromagnetic units, the multiples being chosen so as to make units of convenient sizes for calculations in the technical application of electricity. The electrostatic and electromagnetic units are both "absolute units" that is, are based by definitions on simple relations to the fundamental units, the units of length, mass and time (§150). The particular absolute system long universally used in electricity and magnetism is that based on the centimeter, the gram and second, or the *c.g.s.* system (§150). The following table shows the relations of the practical and absolute electrical units.

ELECTRICAL UNITS

Unit of	Name of practical Unit	Value of practical Unit	
		in C.G.S. E.M.U.	in C.G.S. E.S.U.
Current	Ampere	10^{-1}	3.10^{9}
Quantity	Coulomb	10^{-1}	3.10^{9}
Electromotive force	Volt	10^{8}	$1/(3\times10^{2})$
Resistance	Ohm	10^{9}	$1/(9\times10^{11})$
Capacity	Farad	10^{-9}	9.10^{11}
Inductance	Henry	10^{9}	$1/(9\times10^{11})$

The establishment and universal use of an absolute system of units in electricity and magnetism has contributed much to the progress of the science both in its theory and in its applications. The relations of the units of electric quantity, current, potential, etc., to the units of energy and power are clear and direct in an absolute system. Thus the product of the number of units of current and of potential gives directly the number of units of power or activity, no arbitrary constants entering into the calculations. The advantage of this simplicity is evident. Again the

study of the dimensions of the units (§151), has led to a clearer view of the nature of electrical and magnetic quantities, and of the relations of electrical phenomena to other phenomena. Thus the comparison of the dimensions of the electrostatic and electromagnetic units suggested to Maxwell important similarities of the electrical and optical effects, and contributed much to Maxwell's electromagnetic theory of light (§543). This last theory was again a starting-point for speculations which resulted in Hertz's epoch-making experiments on electric waves and their properties (§542). The dimensions of electrical and magnetic units thus have a greater importance than that of translating results from one absolute system to another (§151).

548. Dimensions of Electrical Units.—The following table gives the dimensions of five of the more usual electrostatic and electromagnetic units.

Name	Symbol	Electrostatic	Electromagnetic
Electric quantity	q	$[L^{\frac{3}{2}}T^{-1}M^{\frac{1}{2}}k^{\frac{1}{2}}]$	$[L^{\frac{1}{2}}M^{\frac{1}{2}}\mu^{-\frac{1}{2}}]$
Magnetic quantity	m	$[L^{\frac{1}{2}}M^{\frac{1}{2}}k^{-\frac{1}{2}}]$	$[L^{\frac{3}{2}}T^{-1}M^{\frac{1}{2}}\mu^{\frac{1}{2}}]$
Magnetic field	H	$[L^{\frac{1}{2}}T^{-2}M^{\frac{1}{2}}k^{\frac{1}{2}}]$	$[L^{-\frac{1}{2}}T^{-1}M^{\frac{1}{2}}\mu^{-\frac{1}{2}}]$
Current	I	$[L^{\frac{3}{2}}T^{-2}M^{\frac{1}{2}}k^{\frac{1}{2}}]$	$[L^{\frac{1}{2}}T^{-1}M^{\frac{1}{2}}\mu^{-\frac{1}{2}}]$
Potential or Electromotive force	$\left\{\begin{matrix} V \\ E \end{matrix}\right.$	$[L^{\frac{1}{2}}T^{-1}M^{\frac{1}{2}}k^{-\frac{1}{2}}]$	$[L^{\frac{3}{2}}T^{-2}M^{\frac{1}{2}}\mu^{\frac{1}{2}}]$

The method of deriving the above dimensions from the definitions is shown by the following examples.

Electrostatic Unit of Quantity. We have by definition (§401) $q=r\sqrt{Fk}$ Using the dimensions of r and F, we get $[q]=[L^{\frac{3}{2}}T^{-1}M^{\frac{1}{2}}k^{\frac{1}{2}}]$. In this k is the specific inductive capacity or dielectric constant (§401), a quantity arbitrarily assumed as unity for air but of undetermined dimensions.

Electrostatic Unit of Current. By definition (§429) $[I]=q/t$. Substituting the dimensions, we get $[I]=[L^{\frac{3}{2}}T^{-2}M^{\frac{1}{2}}k^{\frac{1}{2}}]$.

The starting-point in the electromagnetic system is the definition of unit magnetic pole (§372), $m=r\sqrt{F\mu}$, where μ is the magnetic permeability (§491), a quantity arbitrarily assumed as unity for air, but of undetermined dimensions. From this we get the dimensions of $[m]=[L^{\frac{3}{2}}T^{-1}M^{\frac{1}{2}}\mu^{\frac{1}{2}}]$. From the relation that F, the force at a point in a magnetic field is mH, we get $H=F/m$. The dimensional equation for intensity of magnetic field is thus $[H]=[L^{-\frac{1}{2}}T^{-1}M_{\frac{1}{2}}\mu^{-\frac{1}{2}}]$.

Electromagnetic Unit of Current. The strength of magnetic field at the center of a circular coil of radius r, and carrying a current I, is $H=2\pi I/r$ (§428); substituting dimensions, we get $[I]=[H]\,[L]=[L^{\frac{1}{2}}T^{-1}M^{\frac{1}{2}}\mu^{-\frac{1}{2}}]$.

Electromagnetic Unit of Quantity. From the relation $q=It$, we get $[q]=[L^{\frac{1}{2}}M^{\frac{1}{2}}\mu^{-\frac{1}{2}}]$.

Comparing the electrostatic and electromagnetic units of quantity, we get the ratio $[L^{\frac{3}{2}}T^{-1}M^{\frac{1}{2}}k^{\frac{1}{2}}]\div[L^{\frac{1}{2}}M^{\frac{1}{2}}\mu^{-\frac{1}{2}}]=[LT^{-1}k^{\frac{1}{2}}\mu^{\frac{1}{2}}]$. But LT^{-1} is a velocity (§150). This velocity "v" also appears in the ratios of the other units, though not always as the first power. If an electric quantity is measured in air both electrostatically and electromagnetically, then both k and μ are assumed as unity and the value of this velocity "v" can be determined. This was first done by Weber and Kohlrausch in 1856, by determining the electric quantity in a condenser from its electrostatic capacity and potential (§410), and also by discharging the same quantity through a ballistic galvanometer (§439). They obtained the value $v=310{,}704{,}000$ meters per second. This number is within limits of error the same as the velocity of light. This equality has been established by numbers of later determinations. The close connection between the velocity of light and the ratio of the electrostatic and electromagnetic units confirmed Maxwell in the theory that light is a phenomenon of the same nature as that of electromagnetic actions (§543).

If we assume the equality of the two units of quantity, without assuming k and μ as unity, we get directly that "v"$=1/\sqrt{k\mu}$. This is a very significant relation, and has been the subject of much experiment but has been only partially confirmed.

References

GILBERT, SIR WILLIAM. "*De Magnete*," translation by P. F. Mottelay.
FRANKLIN, BENJAMIN. *Letters on Electricity.*
FARADAY, MICHAEL. *Experimental Researches in Electricity.*
MAXWELL, JAMES CLERK. *Treatise on Electricity and Magnetism.*
HERTZ, HEINRICH. *Electric Waves,* translation by Jones.
THOMSON, J. J. *Conduction of Electricity through Gases.*
The above are the electrical "classics." The present interest in the writings of Gilbert and Franklin is principally historical. Faraday's papers are invaluable in giving an insight into the methods of thought and work of the greatest experimenter in electricity. An inexpensive edition of part of Faraday's "Researches in Electricity" is now available. Maxwell's Electricity and Magnetism is a work to be classed with Newton's Principia as bringing an epoch in science; it uses very advanced mathematical methods. Hertz's and J. J. Thomson's papers are the classics to the two great lines of recent electrical research.
FOSTER AND PORTER. *Electricity and Magnetism.*
STARLING, S. G. *Electricity and Magnetism.*
POYNTING AND THOMPSON. *Electricity and Magnetism.*

BROOKS AND POYSER. *Magnetism and Electricity.*

HADLEY. *Magnetism and Electricity for Students.*

RICHARDSON, S. S. *Magnetism and Electricity.*

THOMPSON, S. P. *Electricity and Magnetism.*

Excellent text-books covering the same subjects, the last four being more elementary than the first three.

JEANS. *Magnetism and Electricity.*

THOMSON, J. J. *Mathematical Theory of Electricity and Magnetism.*

Two good books giving an introduction to the mathematical theory of electricity.

EWING. *Magnetic Induction in Iron and Other Metals.*

The best book on experimental magnetism.

CAMPBELL, N. R. *Modern Electrical Theory.*

CAMPBELL, N. R. *Principles of Electricity.* (People's Library).

FOURNIER d'ABBE E. E. *The Electron Theory.*

DE TUNZELMANN, G. W. *Electrical Theory and the Problem of the Universe.*

These books give interesting discussions of the explanation of electrical phenomena on the electron theory. They are largely or wholly non-mathematical.

LODGE, OLIVER J. *Modern Views of Electricity.*

A very suggestive book by a leader in modern physics.

FLEMING, J. A. *Principles of Electric Wave Telegraphy.*

FLEMING, J. A. *Radio-telegraphy and Radio-telephony.*

FLEMING, J. A. *Waves in Water, Air and Ether.*

The first book is our most complete treatise on wireless telegraphy, and gives the advanced mathematical discussions along with experimental facts. The second book presents the same subject in much shorter form, and is not mathematical. The third book is based on a series of popular scientific lectures on electric waves.

PIERCE, G. W. *Wireless Telegraphy.*

STANLEY, R. *Wireless Telegraphy.*

Excellent books presenting the subject as completely as possible without advanced mathematics.

CARHART, H. S. *Primary Batteries.*

MORSE, H. W. *Storage Batteries.*

LE BLANC. *Electrochemistry.*

NERNST, W. *Theoretical Chemistry.*

LODGE, OLIVER J. *Lightning Conductors and Lightning Guards.*

The most satisfactory book on an obscure subject.

THOMPSON, S. P. *Dynamo-electric Machinery.*

Bulky and containing many details of only special interest, but also giving an excellent presentation of the physics of dynamos and motors.

FLEMING, J. A. *The Alternate Current Transformer.*

A clear presentation of electromagnetic induction in the transformer.

ENCYCLOPEDIA BRITANNICA, eleventh edition.

The electrical articles are by masters and form in themselves a very valuable and complete reference library in electricity and magnetism.

Problems

1. Find the intensity of field at a point 40 cm. from a magnet in the perpendicular bisector of the line joining the poles of the magnet 6 cm. long and of pole strength 160 e.m.u. Calculate the force on a pole of +80 e. m. u. if placed at the point. *Ans.* .0148 e.m. u.; 1.19 dynes.

Magnetic Field.

2. A magnet *NS* 30 cm. long is held vertically; each pole has a strength of 9 units; what is the force on a unit pole at a point 20 cm. horizontal distance from the upper pole? What is the horizontal component of this force?
3. The point *P* is on the perpendicular bisector of a magnet *NS*, at a distance of 30 cm. from *NS*. A pole of strength 8 at *P* is acted on with a force of 3 dynes. Find the moment of the magnet *NS*.
4. To hold a magnetic needle *NS* at an angle of 60° with the earth's field requires a torque of 0.6 dynes acting at a lever arm of 2 cm.; the horizontal intensity of the earth's field is 0.2; what is the moment of the magnet?
5. A short bar magnet is placed with its axis perpendicular to the magnetic meridian, and with the line of the axis passing through the center of a compass needle. At a station X, the compass needle is deflected through an angle ϕ, when the center of the magnet is 40 cm. from the center of the needle. At a station *Y*, the distance from magnet to needle is 35 cm. for the same deflection ϕ. Compare the horizontal intensities.
6. The horizontal intensity of the earth's field at Indianapolis is 0.2, and at Minneapolis it is 0.18; if a magnetic needle makes 100 vibrations at Indianapolis, what will its period be at Minneapolis?
7. To deflect a suspended magnet through an angle of 20° from the magnetic meridian requires 180° of torsion in the wire suspension; how many degrees of torsion must be given the suspension to produce a deflection of 45° from the magnetic meridian?
8. A horizontal magnetic needle makes 40 oscillations per minute at a place where the dip is 70°, and 50 oscillations per minute where the dip is 60°. The total intensity at the first place is 0.6; what is it at the second place?
9. The center of a short bar magnet is at the corner *A* of a square *ABCD* and its axis is in line with the side *AB*. The moment of the magnet is 400, and the length of one side of the square is 60 cm.; find the intensity of the magnetic field at the corners *B* and *D*, due to the bar magnet.
10. Two small spheres, each weighing 1 decigram, having equal charges, are suspended from the same point by silk fibers 80 cm. long. If the spheres are kept 8 cm. apart by repulsion, what is the charge on each?

Electrostatic Fields.

11. Two charges +90 and −40 are 30 cm. apart. Find the intensity of field at a point in the line joining them 60 cm. from the negative and 90 cm. from the positive charge, and calculate the force on a charge of +20 if placed at this point.

12. Two small charged spheres repel each other with a force of 10 dynes when 2 cm. apart. If the charge on one of the spheres is doubled, and the distance between the spheres is doubled, what is the repulsion?
13. What work is done in carrying a charge of 10 units from a point where the potential is 25 to a point where it is 40?

Capacity. 14. A Leyden jar of capacity 10 e.s.u. is raised from a potential -10 e.s.u. to a potential $+15$ e.s.u. Calculate the work required.
15. Given two spheres of radii, 3 cm. and 8 cm., how will a charge of 66 units distribute itself over them if they are connected by a fine wire?
16. What is the charge on a spherical drop of water 2 mm. in diameter, where the electric potential is 100? Two such charged drops unite to form a single spherical drop; assuming no charge is lost, what is the potential of the resulting drop? If three drops thus unite, what is the final potential?
17. A Leyden jar 1/4 cm. thick is 3 cm. in radius and 9 cm. high. Find its capacity, if the dielectric constant for glass is 6. Find charge on each plate when *p. d.* is 15 e.s.u.
18. A condenser of 10 plates, each 20 cm.$\times$30 cm. has 0.4 mm. of air between each pair of plates. Find the capacity.
19. Two plate condensers are joined in parallel. One is a 15 plate air condenser, each plate 11 cm. long and 5 cm. broad, 3 mm. apart; the other a mica condenser of 10 plates, 22 cm. long, 15 cm. broad, 0.5 mm. apart, specific inductive capacity of mica being 8. Find the capacity.
20. Two concentric spheres of radii 10 cm. and 10.3 cm. are separated by air and are charged to difference of potential of 50 volts. Find charge.
21. A pair of circular plates of radii 10 cm. each are 2 mm. apart in air. They are charged to a difference of potential of 20 and are then connected to the plates of an uncharged condenser and the difference of potential falls to 3. Find the capacity of this condenser.
22. Find the capacity of a plate condenser made of two rectangular conductors 32 cm. long and 22 cm. broad, 0.2 cm. apart in air.
23. If the air be replaced by 0.2 cm. sheet of glass of dielectric constant 7, find the charge on each plate when the difference of potential is 20 e.s.u.
24. Find the work in ergs required to charge an insulated metal ball of radius 5 cm. with 20 e.s.u. of electricity.

Magnetic Fields of Currents. 25. A circular coil of 30 cm. diameter has 20 turns. Compute the intensity of the magnetic field at the center when a current of 10 amperes flows through the coil.
26. Find the field strength 16 cm. from the center of a coil in the line of its axis if the coil carry 0.5 amp. and be 24 cm. in diameter.
27. Find force on a pole of 30 e.m.u. if placed at center of coil in problem 26.
28. Calculate current which will deflect a tangent galvanometer 45°, if the galvanometer consists of a coil 18 cm. in diameter, of 7 turns of wire, set up in a field of 0.198 lines per cm^2.

29. The coil of a tangent galvanometer is 34 cm. in diameter and carries a current of 15 amperes; what is the torque on a needle of moment 1.5 at the center?
30. A coil of a tangent galvanometer is to have 10 turns; what should the radius of the coil be, so that the tangent of the angle of deflection of the needle gives directly the current in amperes, at a station where the intensity of the earth's field is 0.19?
31. A circular coil was placed at right angles to the magnetic meridian. The number of oscillations of a small magnetic needle at the center was counted (*a*) when there was no current in the coil; (*b*) when a current i_1 was sent through the coil; (*c*) when a current i_2 was used. For (*a*) there were 40 oscillations per minute; for (*b*) 30 oscillations per minute; and for (*c*) 20 oscillations per minute; what was the relative strengths of the currents i_1 and i_2?
32. A slender solenoid has a length of 50 cm. and has 300 turns of wire; what is the field at the center when the current in the coil is 7 amperes?
33. Calculate the current which produces a magnetic field in the middle of a slender solenoid equal to the earth's field of 0.6, the solenoid being 80 cm. long, and having 400 turns.

Work and Ohm's Law.

34. A current of 6 amperes flows for 4 min. in a circuit of 12 ohms resistance; what is the e.m.f. required? What is the total work done in ergs, also in joules? What is the activity or power in watts?
35. Three electromagnets of resistances 50, 76 and 11 ohms respectively are joined in multiple arc, and a total current of 2 amperes flows through the three; what is the current in each?
36. The total resistance of a circuit is 80 ohms, and on introducing an addition wire the resistance is 66 ohms; what is the resistance of the added wire?
37. A circuit has three branches of 50, 30 and 10 ohms. A fourth branch is put in so that the total resistance is 2 ohms. What is the resistance of the fourth branch?
38. A generator delivers 100 amperes at 110 volts; what is the power in kilowatts and what in H. P.?
39. The resistance of a galvanometer is 126 ohms, and a shunt of 14 ohms is put in; what is the resistance of the shunted galvanometer?
40. By experimenting with a Weston ammeter it was found that 0.00013 amperes through the coil gave one unit scale-deflection. If the resistance of the coil circuit be 5.60 ohms, what must be that of the shunt so that 1 ampere in the external circuit will give 1 unit scale deflection?
41. It is desired to supply 600 incandescent lamps, in parallel, with 1/2 amp. each, at 110 volts potential difference between the lamp terminals. If the drop in the line be 2.2 volts what is the resistance of the line and how much power is lost in it? How much power must be generated and what voltage?
42. A car is lighted by five lamps of 220 ohms resistance each, joined in series. What is the total resistance of the lamps? If the difference

of potential between the ends of the lamp circuit be 550 volts, what current flows through the lamps? What power is expended in this circuit and at 9 c. per kilowatt hr., what does it cost to light a car for one hour?

43. If the motive circuit of a snow sweeper take 50 amp. (at 550 volts) and the broom motors take 80 amp., find the total power consumed in the car if the two circuits be in parallel across 550 volt mains. Find cost per hour at 9 c. per kilowatt hr.

44. A single lighting circuit carries 3 groups of lamps in multiple, the groups being 100 feet apart, and the nearest group being 500 feet from the generator. Each group takes 5 amperes, and the resistance of the line 0.1 ohm per 1000 feet. The potential of the generator is 112 volts. Find the potential of each group of lamps.

45. The distance from power house to library is 650 meters, there are 200 55-watt lamps in library. The e.m.f. at dynamo is 465 volts. Allowing 5 per cent. drop of potential in the line, what must be the resistance of the line. Taking the specific resistance of copper as 1.5×10^{-6} ohms per cm./cm.^{2}, what is the cross-section of the wire? Calculate the watts on full load for this circuit. What must the H. P. of the engine be to carry this?

46. Find the resistance in legal ohms of a tube of mercury at 0°C., 1 meter long, and 1 cm. in diameter.

47. The resistance of a certain firm's copper wire 1 foot long and a mill (one thousandth of an inch) in diameter is 10.7 ohms. What is its specific resistance in ohms per cm./cm.^{2}? 1 inch $=2.540$ cm.

48. The specific resistance of copper is 1.5×10^{-6} and aluminum is 3.2×10^{-6}; with copper at 18 c. per pound, what must be the price of aluminum to compete as an electrical conductor?

49. Given 3 cells of 1.4 volts and 0.8 ohms resistance each, find resistance of the battery if the cells be connected in series and calculate the current through an external resistance of 9 ohms.

50. Find the resistance of the above battery if cells be in parallel and also the current when the external resistance is 9 ohms.

51. Given 20 cells, each with an e.m.f. of 1.7 volts, and an internal resistance of 3 ohms. Calculate the current in the following cases:
   (a) External resistance $R=100$ ohms, cells in series.
   (b) $R=100$ ohms, cells in parallel.
   (c) $R=20$ ohms, cells in series; also in parallel.
   (d) $R=20$ ohms, 4 parallel rows of 5 cells in series.
   (e) Arrangement for maximum current through 20 ohms.

52. A "milli-ammeter," which is to be used as a voltmeter, indicates .005 amperes for a scale division, and has a resistance of 40 ohms. There are 50 scale divisions. What resistance in series with the instrument will enable it to be used for measurements up to 300 volts.

53. It is required to generate 1000 calories of heat per minute in a circuit, the e.m.f. at the terminals of the circuit being 110 volts; what resistance must the circuit have?

Joule's Law. 54. The same current flows through a platinum wire 25 cm. long and 0.5 mm. diameter, and through a copper wire 500 cm. long and 0.6 mm. diameter. What are the relative heat quantities developed in these wires? (Sp. R. of Cu $=1.5\times10^{-6}$. Sp. R. of Pt $=8.9\times10^{-6}$).

55. A uniform current flows for 10 minutes and deposits 4 grams of silver; calculate the current.

56. How much copper can a dynamo giving 30 amperes deposit in an hour?

Electrolysis. 57. How many cubic centimeters of hydrogen at 0°C. and 76 cm. Hg. pressure can a current of 30 amperes produce by the decomposition of acidulated water in an hour? (Sp. gr. of H $=00008.9$ at 0°C. and 76 cm. Hg. pressure.)

Magnetic Induction. 58. An iron anchor ring has 20 cm. mean diameter, and a cross-section of 18 sq. cm. The coil has 600 windings and carries 10 amperes. How great is the magnetic induction B? How many magnetic lines are produced? (The permeability $\mu=200$.)

Electromagnetic Induction. 59. Show that Lenz's law and Fleming's rule lead to the same direction for an induced current in a conductor moved across a magnetic field.

60. The diameter of a circular coil is 30 cm. and the resistance is 0.1 ohm. Find the quantity of electricity in coulombs which will flow in the ring when revolved from a position at right angles to a magnetic field to a position parallel to the field. $H=20$.

61. A circular coil 40 cm. in diameter and with 100 turns is rotated five times per second about a vertical diameter as axis. Find the maximum e.m.f. induced. The horizontal component of the field is 0.2.

62. If the angle of dip is 70°, what is the maximum e.m.f. induced when the above coil is rotated about a horizontal axis parallel to the horizontal component of the field ten times per second?

63. Calculate the e.m.f. induced in a car axle length 120 cm. and with a horizontal linear velocity of 25 meters per second, where the total intensity of the field is 0.6 and the angle of dip is 70°.

64. Calculate the number of revolutions per second which must be given to a disk of 60 cm. diameter to produce an e.m.f. of 5 volts between the center and the periphery of the disk, the axis of the disk being parallel to the field, and the field being uniform and of strength 10,000.

65. A copper disk 10 cm. in radius rotates about a vertical axis with 2000 r. p. m. Given the horizontal component of the earth's magnetic field as 0.2, and the dip as 70°, find the e.m.f. in volts between the center and edge of the disk.

66. Draw a figure showing the directions of the induced currents in the disk of a pendulum swinging between the poles of a magnet across the field. How should the disk be laminated to make the induced currents a minimum?

67. A rectangular coil 10 cm. $\times$ 12 cm. can rotate about a vertical axis which bisects the 12 cm. sides. A current of 3 amperes flows through the coil, and the horizontal intensity of the magnetic field is 0.2. What is the torque (moment of force), when the coil is at right angles to the field?

CONDUCTION OF ELECTRICITY THROUGH GASES AND RADIO-ACTIVITY

By R. K. McClung, D.Sc., F. R. S. C.

Assistant Professor of Physics in the University of Manitoba

CONDUCTION OF ELECTRICITY THROUGH GASES

549. Introduction.—Air, as well as other gases, under normal conditions is almost a perfect non-conductor of electricity. When a difference of potential is established between two points in a gas the gas is in a state of strain, as has been explained in a former paragraph (§398). This strain increases with increase of potential until, when a certain potential is reached, the air is no longer able to withstand the strain and breaks down and a discharge passes. A momentary current of electricity is thus produced through the gas. To produce such a discharge a comparatively large potential is required, several thousand volts being necessary to produce a spark of 1 cm. length in air at atmospheric pressure. The potential necessary to produce a discharge depends upon the shape of the electrodes and the nature and pressure of the gas.

550. Effect of Pressure of a Gas on the Discharge.—If two metal electrodes are inserted in the ends of an air-tight glass tube, such as shown in Fig. 439, filled with air at atmospheric pressure, and if sufficient voltage is applied to the electrodes the discharge ordinarily obtained in air will be observed. If the air be gradually exhausted from the tube the discharge will pass with greater and greater ease as the pressure is diminished, until a certain minimum pressure is reached, and if the exhaustion be carried beyond this point the voltage necessary to produce a discharge will increase somewhat rapidly, until at the lowest pressure obtainable it will be impossible to cause a discharge to pass at all.

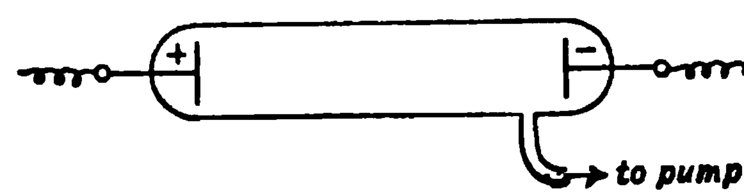


Fig. 439.

The pressure corresponding to this minimum potential is called the critical pressure and varies with the distance between the electrodes.

As the pressure is gradually diminished below atmospheric pressure the appearance of the discharge changes very much. At first the spark becomes more regular and uniform between the electrodes, then broadens out and assumes a fuzzy appearance of a bluish color. When a pressure of about half a millimeter is reached the discharge assumes a very marked appearance, which is shown in Fig. 440. The surface of the negative electrode or

FIG. 440.

cathode is covered by a very thin layer of luminosity; next to this is a dark space which is called the *Crookes dark space;* immediately beyond this dark space is a luminous part called the *negative glow*, and then beyond this again is a second dark region, sometimes called the *Faraday dark space.* Between this and the anode there is a luminous region which goes under the name of the *positive column.* Under certain conditions of current and pressure the positive column shows alternately dark and light spaces which are called *striæ.* The proportion of the space between the electrodes occupied by each of these sections of the discharge depends upon the distance between the electrodes. Any increase in this distance beyond a few centimeters causes an increase in the length of the positive column but no increase in the negative glow or dark space. Similar phenomena occur in other gases besides air.

551. Cathode Rays.—When the pressure in such a discharge tube is lowered to the neighborhood of a hundredth of a millimeter, a new phenomenon makes its appearance. The positive column begins to disappear and a bright phosphorescence appears on the sides of the tube. This phosphorescence appears to be produced by radiations or streams of very minute particles issuing normally in straight lines from the cathode. They are, consequently, called *cathode rays*, and possess remarkable properties.

If a magnet is brought close to the tube the rays are deflected

from their original path. A solid body placed inside the tube in the path of the rays casts a well-defined shadow. If the rays be concentrated upon a solid body inside the tube, such as a platinum plate, it may be heated even to incandescence.

One of the most important properties of the cathode rays is that they *carry a negative charge of electricity*. This was originally proved by Perrin and his method was later modified by J. J. Thomson. A diagram of the apparatus used in the latter experiment is shown in Fig. 441. *A* was the cathode and *B* the anode. The cathode rays from *A* passed into the larger part of the tube through a hole in *B* and fell upon the glass at a point *C*. A side tube contained two coaxial metal tubes. The outer one *E* had a slit in the end and was connected to earth. This shielded the inner tube from any stray electric effects. The inner tube *D* had a slit opposite that in *E* and was insulated from *E* and connected to an electrometer. When the cathode rays were allowed to fall upon the glass bulb the electrometer indicated only a very small

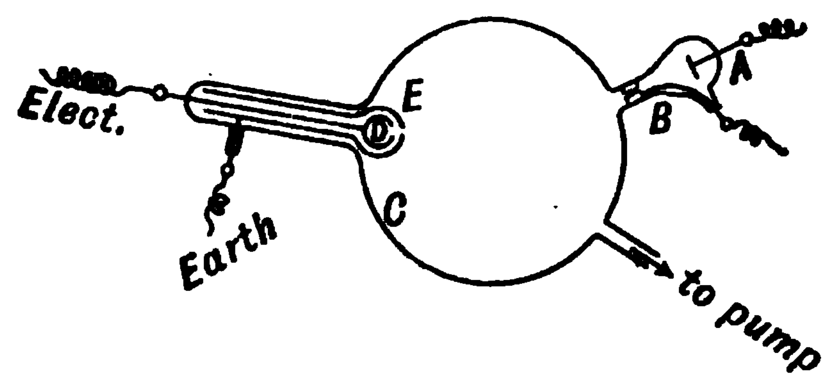


FIG. 441.

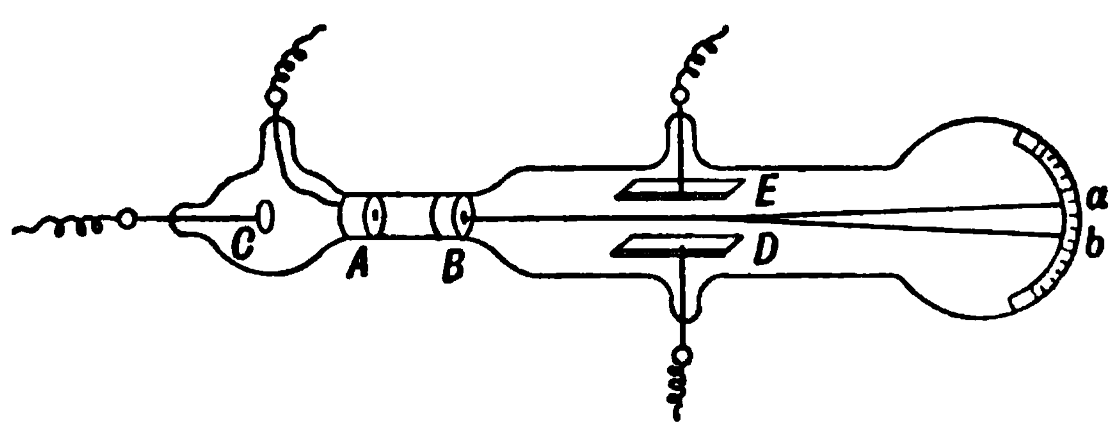


FIG. 442.—(After J. J. Thomson, Conduction of Electricity through Gases.)

effect, but if the rays were deflected by means of a magnet so that they fell upon the slits in the cylinders *D* and *E* the electrometer indicated that the cylinder *D* had received a considerable negative charge. If the rays were deflected still further so as to miss the slit the cylinder immediately ceased to receive any charge. This experiment clearly shows that the rays are accompanied by a negative charge of electricity. If the cathode rays be allowed to pass between two parallel plates inside a highly exhausted cathode ray tube, such as is indicated by Fig. 442, and a large

difference of potential be established between the plates, the beam of rays will be deflected and the deflection will be in the same direction as a negatively charged particle would be moved by the field.

552. Velocity, and Ratio of the Charge to the Mass, of a Cathode Ray Particle.—We will now consider the method which J. J. Thomson originally used to determine experimentally the velocity of these particles and the relation between the mass of a particle and the charge which it carries. A highly exhausted cathode ray tube was arranged as shown in Fig. 442. C was the cathode, A the anode, and B a thick metal plug. A and B were pierced by holes in the same straight line about a millimeter in diameter, so that a very narrow beam of rays might pass along the middle of the tube and fall upon a screen of phosphorescent material, thereby producing a small bright spot. D and E were two parallel plates which could be connected to the poles of a battery. Suppose that V is the velocity of the particle in cms. per sec., m its mass, and e the charge which it carries, measured in electromagnetic units. If the tube be placed between the poles of a strong electro-magnet, so that a field of strength H is acting at right angles to the beam, the spot on the screen will move from a to b in a direction at right angles to the lines of force. The cathode particle will follow a curved path just as a moving projectile follows a curved path when acted on by gravity. Let the radius of curvature of this path be r. The deflecting force acting along this radius of curvature is proportional to the magnetic field, the charge on the particle, and its velocity and is, consequently, equal to HeV (see §433). This force must equal the centrifugal force of the particle which, from dynamics, is equal to mV^2/r (see §§32, 47). Therefore,

$$HeV = \frac{mV^2}{r}$$

$$\therefore\ Hr = \frac{mV}{e} \qquad (1)$$

H can be measured and r may be found from ab and the dimensions of the apparatus. Therefore the quantity mV/e is known. Suppose now that a difference of potential be established between the plates; an electric force will act on the beam of rays and if it is applied in the right direction it will tend to deflect the

beam in a direction opposite to the magnetic deflection. Let the magnetic and electric forces be adjusted so that their effects on the particles exactly balance each other, then the phosphorescent spot will return to the position it had before any force acted on it. Let this electric field be X. The force acting on the particle will then be Xe and, therefore, if the electric and magnetic forces exactly balance each other

$$Xe = HeV$$
$$\therefore V = X/H \tag{2}$$

X and H can both be measured and, therefore, V may be determined, and knowing V the value of e/m is easily found from equation (1).

By this method Thomson found the value of V to be 2.8×10^9 cms. per second, which is just about one tenth the velocity of light. This value is not quite constant as it varies somewhat with the potential in the tube. He also found a value for e/m the magnitude of which, according to later determinations, is 1.7×10^7; and he discovered that it was independent of the nature of the gas in the tube.

The greatest value of e/m known in electrolysis is found in the case of the hydrogen ion and is about 10^4. The value for the cathode ray particle is thus 1700 times that for the hydrogen ion. In a later paragraph (§564) the charge e carried by the cathode particle will be determined and it may be shown to be the same as for the hydrogen ion. Consequently, the mass of the cathode particle must be about 1/1700 of the mass of the hydrogen ion or atom. This cathode particle possesses the smallest mass yet known, and is called an *electron* or negative corpuscle.

553. Röntgen Rays.—In 1895 Röntgen observed that some sort of radiation was produced outside an ordinary cathode ray tube. Phosphorescent bodies placed near the tube were strongly affected and a photographic plate in the neighborhood became blackened. These radiations have been called Röntgen rays after their discoverer. The name first applied to them was X rays and this name is still often used.

The method of producing Röntgen rays is shown in Fig. 443. AB is a large glass bulb. The cathode a consists of a concave piece of metal, usually aluminum. The cathode rays proceed

normally from the surface of *a* and on account of its concavity are brought to a focus at the point *c*, and hence the name "focus tube." The anode *b* consists in its simplest form of a flat platinum plate which is placed at an angle of 45° to the axis of *a* and so that the center of curvature of *a* is at the point *c*. The Röntgen rays travel outward in all directions from *b*. To generate the rays the electrodes are connected to the terminals of the secondary of an induction coil or to an electrostatic machine.

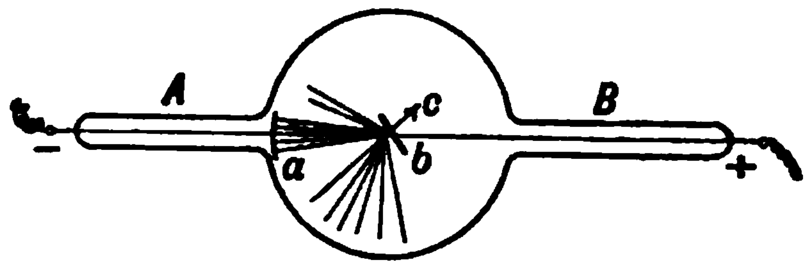


FIG. 443.

Röntgen rays differ from cathode rays inasmuch as they are able to penetrate bodies of considerable thickness. Their penetrating power, as well as some of their other properties, depends upon the conditions existing within the tube from which they originate. With a very low pressure within the tube and, consequently, a large potential difference between the electrodes, the rays produced are very penetrating, being capable of going through several inches of wood and even several millimeters of lead. Such rays are usually called "hard rays." In the case of a higher pressure and smaller difference of potential the rays are less penetrating and are called "soft rays." Different substances absorb the rays of any particular type to a different degree. Generally speaking dense substances produce greater absorption. It is this variation in the absorptive power of substances which enables us to make Röntgen ray photographs. Röntgen rays act upon a photographic plate in a manner similar to ordinary light and the effect produced depends upon the intensity of the rays. Thus a photograph of the bones of any portion of the human body may be obtained, for bones being denser than the flesh absorb the rays more and, consequently, the intensity of the rays which have traversed the bones is less than the intensity of those which have passed through only the flesh.

The Röntgen rays travel in straight lines with very high velocity. Marx has shown that they travel with the velocity of light, that is, 3×10^{10} cms. per sec. No evidence has as yet been found of any refraction of the rays when they pass from one medium to another, nor has it so far been possible to deflect the rays by a magnetic field.

It has been shown mathematically that when an electrically charged particle is suddenly brought to rest an electromagnetic disturbance is produced in the surrounding medium and travels outward. This condition is fulfilled when a cathode ray particle is suddenly arrested by striking against any solid body. All of the evidence is consistent with the view that Röntgen rays are electromagnetic disturbances of the same general nature as light waves.

554. Reflection of *X*-rays by Crystals.—When a narrow beam of *X*-rays falls upon a crystalline substance and, after transmission by the crystal, is allowed to fall upon a photographic plate some very remarkable results are observed. The photograph obtained shows a number of spots arranged in a regular manner and forming a definite pattern. This pattern is explained by the principle of diffraction of the waves or pulses of the *X*-rays caused by reflection from the series of parallel planes of which the crystal is made up. Crystals may be considered as made up of atoms arranged in certain definite planes. If these planes form a series in which they are parallel and equally spaced then a pulse falling upon this series of planes will be reflected from them and form a wave train. By a careful study of these diffraction patterns caused by this reflection from the planes of the crystal a considerable amount of information in regard to crystal structure is obtained. For instance the wave length of a homogeneous beam of *X*-rays can be found in terms of the dimensions of the various parts of the crystal and from this and other data in regard to the mass of the atoms certain dimensions of the crystal can be determined and the wave length of the *X*-ray pulse can be calculated. By this means the various wave lengths of the rays produced by any *X*-ray bulb can be obtained and the *X*-ray spectrum determined. By data of this kind it has been shown that the wave length is characteristic of the anode in the bulb from which the rays are produced.

555. *X*-ray Spectrometer.—For the purpose of studying these *X*-ray spectra Bragg has devised a very ingenious *X*-ray spectrometer. It corresponds in general arrangement to an ordinary spectrometer. A lead screen pierced by a narrow slit through which the *X*-rays pass takes the place of the ordinary collimator; the reflecting crystal occupies the position of the usual prism,

while an ordinary ionization chamber takes the place of the common telescope. The beam of rays from the bulb enter the narrow slit and fall upon the crystal and are reflected into the ionization chamber where the ionization produced by them is measured. The crystal and the ionization chamber can be rotated so that the effects for various angles of incidence and reflection can be measured. By this means the reflection produced at various angles by the crystal can be examined in detail. This method of studying *X*-rays has been a fruitful one in giving information in regard to the nature of the rays and also the structure of crystals.

556. Conductivity of Gases Produced by Röntgen Rays.—If a well-insulated body, such as the leaves of a gold-leaf electroscope *E* (Fig. 444), be charged up in thoroughly dry air the charge will be retained for many hours. If, however, a beam of Röntgen rays passes through the gas surrounding the leaves they will

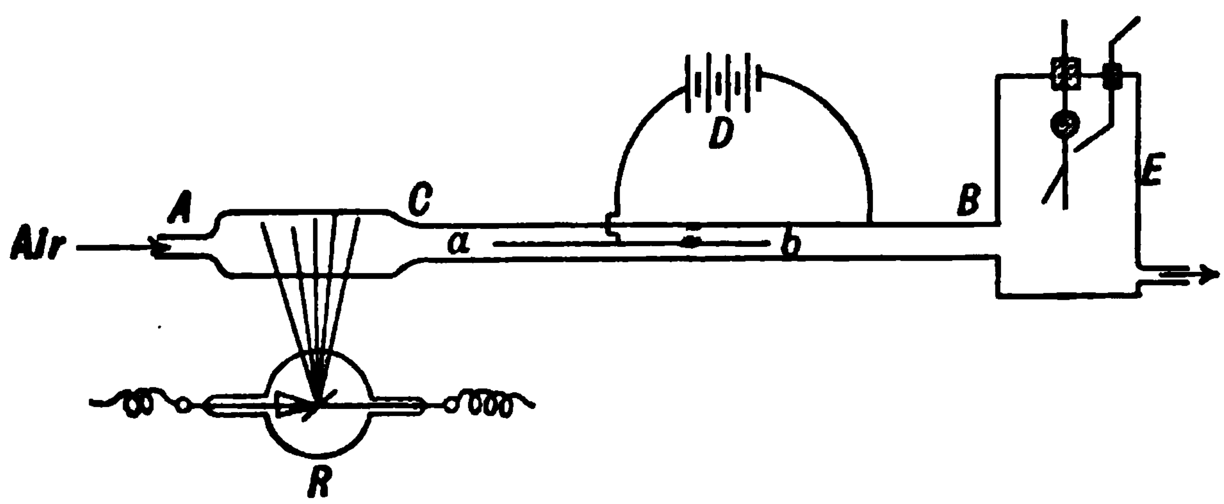


FIG. 444.

immediately lose their charge and collapse, showing that the air must have become conducting, allowing the charge to leak away.

Instead of the rays falling directly upon the gas surrounding the leaves of the electroscope let a system be arranged as shown in Fig. 444. *AB* is a metal tube through which a stream of air may be sent and which leads into an electroscope *E*. If the Röntgen rays fall upon the air in the part *AC*, no effect is produced in the electroscope as long as there is no stream passing through the tube; but as soon as a stream of air is passed through the tube into *E* the leaves lose their charge. This conducting property imparted to the air by the rays, therefore, may be transported along with the air. If a plug of cotton wool be placed in the tube at *C*, or if the air be bubbled through water, after being acted

upon by the rays this conductivity is entirely destroyed. If an insulated wire *ab* be introduced in the center of the tube *CB* and a strong electric field be established between the wire and the tube, by connecting the wire to one pole of a battery and the tube to the other pole, the air loses its conductivity in passing through the tube.

The removal of this conducting power from the gas by filtering it through cotton wool or water indicates that the conductivity must be due to something mixed with the air, while its removal by an electric field shows that, whatever it may be that is mixed with the air, it must carry an electric charge.

Suppose again that *A* and *B*, Fig. 445, are two parallel metal plates placed a few centimeters apart in air and let *A* be connected to one pole of a battery while the other pole is connected to earth; let *B* be connected to one pair of quadrants of a quadrant electrometer while the other pair of quadrants is connected to earth. If a beam of Röntgen rays be passed between these plates it will be observed that *B* immediately begins to receive a charge, as indicated by the deflection of the electrometer needle. It will continue to charge up as long as the rays are acting, but will cease if the rays cease. If C_1 is the positive pole of the battery then *B* will receive a positive charge, but if the poles be reversed *B* will receive a negative charge. The rays thus apparently cause a transference of electricity through the air to *B* and the sign of the electric charge given to *B* depends upon the sign of *A*.

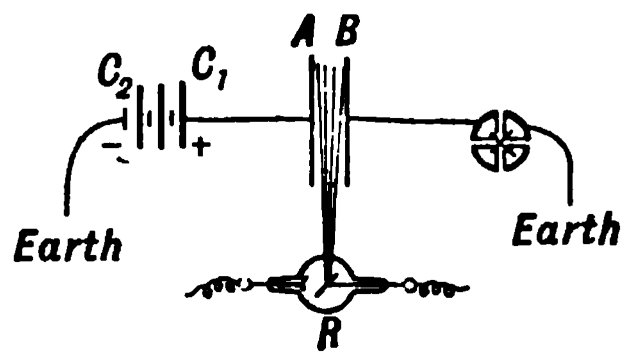


FIG. 445.

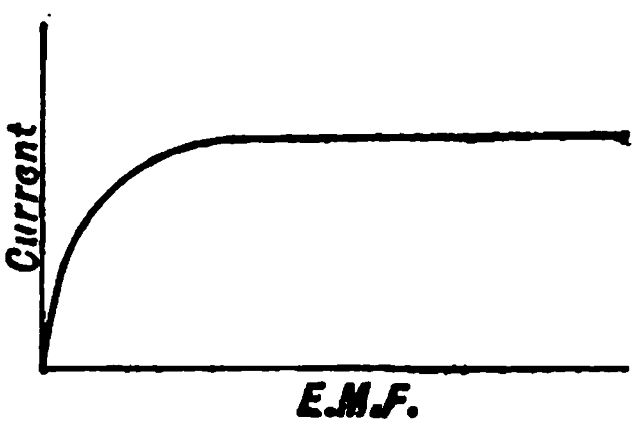


FIG. 446.

557. Saturation Current.—If the potential difference between *A* and *B* be altered the charge received by *B* in a given time alters, that is, the current between *A* and *B* depends upon the voltage. The current through the gas does not, however, obey Ohm's law, for if the current corresponding to different voltages be measured and a curve plotted showing the relation between current and voltage, it will assume the form shown in Fig. 446 instead of being a straight line. It will be seen that for small voltages the current

obeys Ohm's Law, but it soon begins to fall off and finally reaches a constant value even for a large increase in voltage. This characteristic curve has been called a saturation curve on account of its similarity in form to the saturation curve in the magnetization of iron. The current corresponding to the flat part of the curve is called the saturation current.

The current through a gas differs very markedly in another respect from the current through metals or liquids. When the distance between two electrodes immersed in a liquid is increased the current decreases on account of the increase of resistance between the electrodes, but in the case of a gas the saturation current increases when the distance between the plates is increased. Within certain distances the saturation current is proportional to the distance between the plates.

558. Theory of Ionization of Gases.—These facts along with others have led to the ionization theory of gases. According to this theory Röntgen rays, when they pass through a gas, cause the molecules of the gas to be broken up into positively and negatively charged carriers of electricity called ions. This process of breaking up the molecule is called ionization and the gas is said to be ionized. From each molecule ionized two ions, having equal charges but of opposite sign, are produced. The transference of electricity through the gas is due to the movement of these charged carriers under the influence of an electric field. The positive ions are attracted to the negative electrode and the negative ions to the positive electrode, and the movement of these electric charges constitutes a current. When the gas is passed through the tube with a central wire, between which there is an electric field, the positive and negative ions are attracted to the negative and positive electrodes respectively and thus removed. When the gas is passed through cotton wool the ions are caught by the wool.

559. Explanation of Saturation Current.—The above theory also explains the saturation curve for a current between two plates. The current is proportional to the number of ions reaching the plates per second and, therefore, to the potential difference provided this is not too high. But when the voltage reaches a certain value the ions move so fast that they practically all reach the plates before they have time to recombine, and the current could not be increased further even by a higher voltage, as the number of ions removed could not be augmented.

The increase of current between two plates when the distance between them is lengthened is also easily explained by this theory. When the plates are placed farther apart the volume of gas acted on by the rays is increased

and, consequently, the number of ions produced grows greater in the same proportion, and a greater number will reach the plates per second and the maximum current be raised.

560. Effect of Conditions on Ionization.—The nature and quality of the ionizing rays determine the number of ions produced in any given gas. Cathode and Röntgen rays, for instance, differ in ionizing power and even Röntgen rays differ among themselves in this respect. Penetrating rays of any type are usually less powerful ionizers than those less penetrating. For a constant ionizing source the number of ions produced in a given volume of gas is found experimentally to be directly proportional to the pressure. Temperature on the other hand has, as far as is known, no effect on ionization, if the density of the gas is kept constant.

561. Recombination of Ions.—When the rays begin to ionize the gas the ions gradually increase in number until a steady state is reached, when no further increase will take place no matter how long the rays act. As the rays are continually producing ions they must be disappearing at the same rate as they are being produced when this steady state is reached. Being positively and negatively charged bodies and being in motion they collide and neutralize each other electrically and disappear as far as producing any conductivity is concerned.

562. Diffusion of Ions.—The ions of an ionized gas are in motion and if there is an excess of ions in one part of the gas they will diffuse to the other part. If the ionised gas is in an enclosed vessel the ions will diffuse to the sides of the vessel and disappear from the gas. Sometimes, in a very confined space, the loss of ions by diffusion is even more important than the loss by recombination.

A detailed study of diffusion has led to the theory that both the positive and negative ions, at ordinary pressures, consist of a cluster of molecules surrounding a charged nucleus. Ionisation is considered to consist in separating a negative electron from the neutral molecule and then the electron becomes loaded with a cluster of molecules and forms the negative ion under ordinary conditions. The positive ion consists to begin with of the molecule deprived of the electron and then a cluster of molecules is formed about this positively charged center. The positive and negative ions diffuse more nearly at the same rate in moist than in dry gases, for in dry gases the negative ion is smaller, but in a moist gas it becomes more loaded up with moisture than the positive ion and its rate of diffusion decreases more rapidly. As the pressure of the gas is lowered the coefficient of diffusion of the negative ion increases faster than that of the positive and it has been shown that at low pressures the negative ion is the same as the electron.

563. Ionization by Collision.—In §557 the current-voltage curve for a gas at atmospheric pressure showed a final maximum current. In a gas at low pressures, in the neighborhood of 1 mm. of mercury a new phenomenon appears and the corresponding curve for current and voltage assumes the from shown in Fig.

447. For low voltages the part of the curve up to a point A is of the same form as the saturation curve at atmospheric pressure, but when the voltage is increased beyond a certain amount the current begins to increase again, at first slowly and then very rapidly. The increase of current beyond the point A must be caused by an increase in the number of ions due to some cause other than the original ionizing agency. This larger number of ions has been explained by the theory that if an ion is moving with sufficient velocity it will produce more ions by collision with the molecules of the gas. A moving ion possesses kinetic energy and if its velocity is great enough it will possess sufficient energy to ionize a molecule with which it may collide. The kinetic energy depends upon the velocity and this, in turn, depends upon the electric field and upon the opportunity the ion has of acquiring speed among the molecules of the gas. At atmospheric pressure the molecules are so close together that the ion is not able, between two collisions, to acquire sufficient velocity in ordinary electric fields to ionize a molecule, but at low pressures the molecules are so few in number and so far apart that the ion may acquire sufficient speed between collisions to ionize any molecule which it strikes. This production of ions by collision is only observed for ordinary electric fields at pressures below about 30 mm.

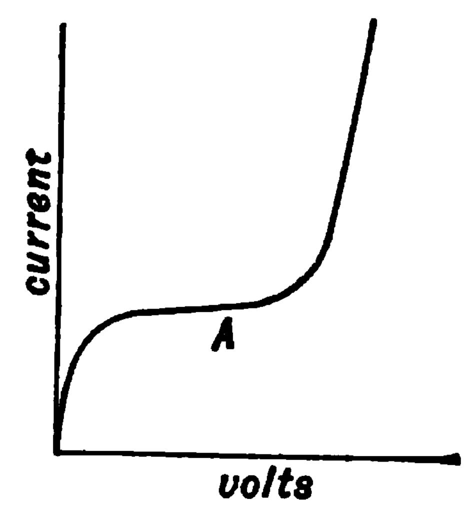


FIG. 447.

The above theory of ionization by collision furnishes a very satisfactory explanation of the electric spark through a gas at atmospheric pressure. There are always in gases a few ions which may be detected by the use of sensitive instruments. If a voltage high enough to produce a spark is established between two points, the few ions naturally present in the field will acquire a velocity sufficient to ionize any molecules against which they strike; these new ions will in turn produce more ions, and so the number will increase very rapidly until there are enough to carry a current, and this current is the electric spark.

564. Charge carried by an Ion.—It was known for some years that if dust particles were present in a damp gas the water vapor would condense around these nuclei when a sudden expansion of

the gas took place. If a beam of X rays is allowed to fall upon a steam jet, condensation takes place, the ions produced acting as nuclei on which water vapor condenses.

This property of ions to act as condensation nuclei has been utilized by J. J. Thomson to determine the absolute value of the charge carried by an ion. When an expansion takes place in ionized air water drops form around the ions and fall under the action of gravity. Sir George Stokes has shown that a drop of water of radius, r, falls through a gas of viscosity, μ, with the velocity, v, given by the equation

$$v=\frac{2}{9}\frac{gr^2}{\mu}$$

where g is the acceleration of gravity. The velocity, v, can be measured by observing the rate at which the cloud falls under the action of gravity, and since g and μ are known r may be determined. If m is the mass of water deposited and n the number of drops per c.c. then $m=n\times\frac{4}{3}\pi r^3$, since the density of water is unity. The amount of water vapor deposited when a known expansion occurs can be easily calculated from well-known thermal considerations and, therefore, m may be determined. Knowing m and r the number of drops, n, which is the same as the number of ions, is easily calculated.

If all the ions present be extracted by an electric field between two electrodes in the usual way, the total charge carried by all the ions can be measured. Knowing, therefore, the number of ions and the total charge on them, the charge carried by each one is determined. By a modification of this method using a single drop instead of the cloud Millikan has shown this charge to be equal to 4.774×10^{-10} electrostatic units. It has been shown that the charge acquired by such a drop suspended in space is always an exact multiple of the elementary charge. The charge carried by ions in hydrogen and oxygen has the same value and does not depend upon the source from which they are produced.

565. Emission of Electrons by Metals.—If ultra-violet light rays fall upon the clean suface of a piece of zinc, sodium, potassium, lithium, etc., which is negatively charged the metal will lose its charge, while if the metal be uncharged to begin with it will acquire a positive charge. If the metal is positively charged to

begin with no loss of charge takes place. These *photo-electric effects* as they are called, have been shown to be due to the liberation of negative corpuscles, or electrons, from the metal by the action of the ultra-violet light.

If a metal electrode be placed near to a metal wire and the latter be then heated until it begins to glow, a current through the gas will be produced and the electrode will receive a charge. A platinum wire heated to redness will under some conditions give a positive charge to the other electrode, but if heated to white heat the charge is negative. The behavior of hot metals is somewhat irregular but in general metals and carbon heated to incandescence in high vacua give off negatively charged carriers. The ratio of the charge to the mass of these carriers has been shown to be the same as for the cathode ray particles and the electron liberated by ultra-violet light at low pressures. This along with other considerations has led to the theory that these negative corpuscles are distributed throughout the volume of metals at all temperatures, but when the metals are heated to incandescence the corpuscles then acquire sufficient energy to escape into the the surrounding space.

566. Ionization by Flames.—If two electrodes are placed some distance apart in an ordinary Bunsen flame quite an appreciable current is observed which may be measured by a galvanometer. If the air surrounding such a flame be drawn away from the flame it is found to be still a conductor. The ions which have been produced in the gas by the flame appear to be much larger than those produced in other ways, for their velocity has been measured and found to be much less than that of other ions. It is due to this conducting power of flames that when an insulator has received an electrostatic charge it may be discharged by simply passing a Bunsen flame over it

RADIO-ACTIVITY

567. Discovery of Radio-Activity.—The phosphorescent action of Röntgen rays led physicists to investigate phosphorescent substances and Becquerel in 1896 found that the double sulphate of uranium and potassium emitted a radiation which produced an effect upon a photographic plate similar to that of X-rays. He later examined other compounds of uranium as well as the element itself and found that they all possessed this power. Although the phosphorescent action of Röntgen rays pointed the

way to this discovery, it has since been shown that there is no connection between the rays emitted by uranium and its phosphorescence, for some compounds which are not phosphorescent emit the rays.

Becquerel and others showed that these radiations from uranium were capable of discharging electrified bodies and that this power of discharging electrified bodies was due to the production by these rays of ions in the gas, similar to the ions produced by Röntgen rays.

If the rays from uranium be allowed to pass between two parallel plates, between which there is a difference of potential, a current will pass through the air just as in the case of a gas ionized by Röntgen rays. Thus suppose that A and B (Fig. 448) are two insulated metal plates. The upper plate A is connected to one pair of quadrants of an electrometer, the other pair being to earth. If a layer of one of the compounds of uranium be sprinkled on the plate B, as indicated, an ionization current will be produced between A and B.

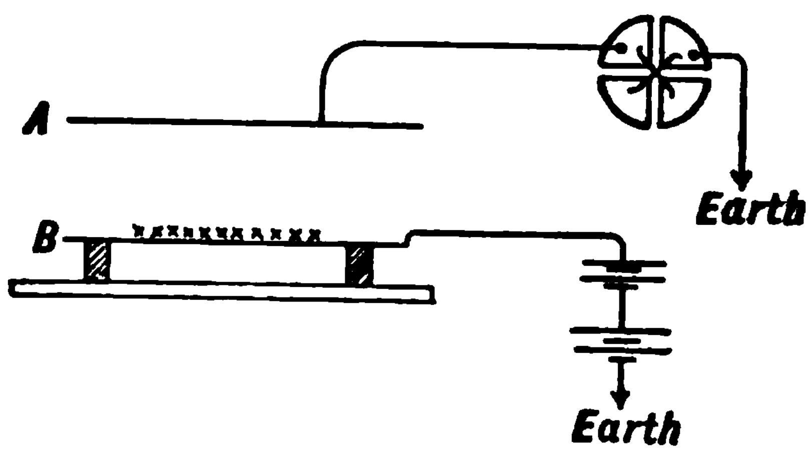


Fig. 448.

This property of uranium does not deteriorate with time. Uranium and other bodies, which possess similar properties, are called *radio-active bodies.*

568. Other Radio-active Substances.—Schmidt, and independently Mme. Curie, discovered that the element *thorium* and its compounds possess radio-active properties. The photographic action of thorium was found to be distinctly weaker than that of uranium, while the ionizing action was about equal to that of uranium, but was very irregular. A very systematic examination of a large number of minerals containing uranium and thorium was then undertaken. Using the electrical method the current produced between two plates by a given amount of each of the minerals was measured. The results showed that all minerals containing uranium or thorium were radio-active, but that several specimens of pitchblende, as well as some other minerals, were several times more active than uranium itself. T is led to the conclusion that there must be some other and

more active substance in pitchblende. M. and Mme. Curie then investigated this question chemically and discovered two new active bodies.

The first of these substances to be separated by purely chemical means was found to be very much more active than uranium and it was given the name *polonium* in honor of Mme. Curie's native country. Polonium differs from uranium in the essential particular that its activity is not constant but gradually dies away. In some cases it was found that at the end of about six months after preparation the activity had fallen to half its original value.

The other active substance discovered in pitchblende was found to be enormously more active than uranium. In its pure state it is about a million times more active, and it was called *radium* by the discoverers. The quantity of radium in pitchblende is almost infinitesimal, about a ton of pitchblende containing only a few milligrams of pure radium. Radium is found in varying quantities in a number of minerals and in various parts of the world. For some years the pitchblende found in Bohemia furnished most of the radium, but quite recently considerable quantities have been obtained from the carnotite ores found in Colorado.

In practice radium is not separated from its compound but is usually employed in the form of the bromide, and what is often called "pure radium" is really "pure radium bromide." It also forms other compounds, such as the chloride, sulphate, etc., and these salts are all naturally phosphorescent and their radiations produce phosphorescence in various substances such as platino-barium cyanide, willemite, etc.

Debierne, in analyzing residues from pitchblende, discovered a very active substance which he called *actinium*. The properties of actinium are very similar to those of thorium, but the former is very many times more active than the latter. Actinium besides being strongly radio-active is capable, like radium, of producing phosphorescence in such substances as zinc sulphide, willemite, etc.

569. Three Types of Rays.—In examining the radiations from uranium Rutherford found that there were two distinct types of rays, one type which were easily absorbed by solid bodies, and a

second type which were more penetrating and, besides, could be easily deflected from their path by a magnetic field. The former he called *α rays* and the latter *β rays*. Later it was shown that there was still a third type emitted which were extremely penetrating and could not be deflected by a magnetic field. These were called *γ rays*. The four radio-active substances uranium, thorium, radium and actinium, under normal conditions, give out these three types of rays. Polonium, however, emits only α rays.

570. The β Rays.—Becquerel, using the photographic method, showed that the β rays of radium behaved in every respect like cathode rays. They, consequently, must be *negatively charged particles or electrons.*

Combining deflections, by a magnetic and by an electric field, in a manner somewhat similar in principle to that used in the case of the cathode rays (§552), Becquerel determined the velocity and the ratio of e/m for the β rays. For e/m he found a value which does not differ much from the value found for the cathode rays or *electrons.* He observed, however, that the β rays did not all have the same velocity as some were bent more than others. He showed that the velocities varied from about 6×10^{9} to 2.8×10^{10} cms. per sec., the latter approaching very nearly the velocity of light which is 3×10^{10} cms. per sec. The β rays from radium appear, therefore, to be complex, being a mixture of rays of the same nature but travelling with different speeds. The β rays from uranium differ from those of radium in this respect for the former appear to be homogeneous.

571. Nature of the mass of an electron.—This complexity of the β rays (or electrons) with regard to velocity led Kaufmann to examine whether the value of e/m for these rays varied with the speed. He showed that e/m decreased when the speed increased. Assuming that the charge on the β ray particle is constant the mass of the particle appears to increase with increase of velocity.

Several mathematical physicists have worked out from purely theoretical considerations that the apparent mass of a moving electron is due, either wholly or in part, to the electric charge in motion, that is, when an electric charge is moving it appears to possess what corresponds to inertia, due to the fact of its being in motion. This apparent inertia according to this view is not due

to material mass as we are accustomed to conceive of it, but is a result of the motion of the electric charge. These theoretical considerations further show that this apparent mass, which seems to be electrical in origin, increases with the speed of the moving charge. Experimental results seem to confirm the theoretical view that *the mass of the electron is due, wholly or in part, to the fact that the electric charge is in motion.*

572. The α Rays.—The first attempts to deflect α rays by a magnetic field and so ascertain their nature failed. Rutherford succeeded in doing this by using intense radiation and a very powerful field. His apparatus is shown diagrammatically in Fig. 449.

A is a gold-leaf electroscope, *SS*, a set of parallel brass plates separated by very narrow slits the width of which was in some experiments as small as 0.042 cm. but varied for different experiments up to 0.1 cm. A quantity of radium, *R*, was placed below the slits and the rays passed up through them and into the electroscope where they ionized the air. Of course, the β and γ rays were also present but the ionization produced by the α rays was more than nine times that produced by the β and γ rays combined, so their presence did not affect the experiment. By applying a magnetic field in a direction parallel to the slits and at right angles to the plane of the paper the rays, if they are deviable, should be bent either to the right or left and strike the plates and be stopped before they could emerge beyond the slits. He found that by the application of the magnetic field over eight-ninths of the α radiation could be cut off, showing that the α rays could be deviated by the field. By a slight modification of the experiment he showed that they were bent in the opposite direction to that in which the β rays would be bent, indicating that the α rays must carry a positive charge. He also succeeded in deflecting the α rays by an electric field using an apparatus similar to the one just described.

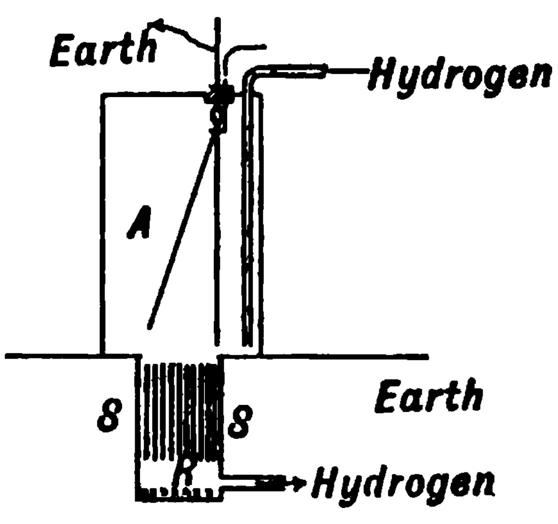


FIG. 449.

Experiments show that within the limits of experimental error the value of e/m is the same for the α rays emitted by the various

radio-active substances. The average experimental value obtained is about 5×10^3 electromagnetic units. Assuming that the charge on each particle is the same, the mass of the α particles emitted by the different substances is constant.

Although the mass is constant yet the velocity of expulsion of the α particles is not the same for all substances, as it is found to vary from 1.56×10^9 to 2.25×10^9 cms. per second.

573. Mass and Nature of α Particle.—These results enable us to obtain a more definite idea of the mass and nature of the α particle. The value of e/m for the atom of hydrogen liberated in the electrolysis of water is about 10^4 electromagnetic units, while we have just seen that for the α particle e/m is 5×10^3. Rutherford showed that within the limits of experimental error the charge carried by the α particle is twice the charge carried by a gaseous ion and consequently twice the charge on the electrolytic hydrogen ion or atom. It follows from this that the mass of the α particle must be four times the mass of the hydrogen atom. Since it is atomic in size and of the same order as the atom of helium (whose atomic mass is 3.96 in terms of hydrogen) and since there does not seem to be any place according to the periodic law among the elements for a new one in that part of the series the most natural hypothesis is that *the α particle is an atom of helium carrying twice the ionic charge of hydrogen.* Helium is continually produced by both radium and actinium. As a final proof it has been shown that when α particles are allowed to penetrate into a vacuum helium always accumulates.

574. Absorption of α Rays.—A distinguishing characteristic of the α rays is that they are very easily absorbed when passing through either gases or solids. The proportion of the rays absorbed by a given thickness of any solid may be determined by first measuring the saturation current produced by the rays, and then covering the radiating material with the absorbing solid and again measuring the current produced by the rays after they have passed through the solid. The absorbing layer must be very thin or else all the rays will be stopped. The most penetrating α rays known are completely absorbed by a thickness of only about 0.006 cm. of aluminum. The penetrating power of the α rays varies greatly with the different substances from which they are emitted.

The α rays are very easily stopped by gases, a few centimeters of air at atmospheric pressure being sufficient to absorb them, consequently, the ionisation produced by them exists only within a few centimeters of the source from which the rays come. The absorption by gases depends upon

the density, being in some cases proportional thereto, but not so in all cases. The absorption of the rays by gases is important as the degree of ionisation produced by the rays depends upon the amount of the rays absorbed, the relative ionisation by the α rays in gases being directly proportional to the relative absorption.

575. The γ Rays.—The third distinct type of rays given out by some of the radio-active substances differs very essentially from the α and β rays. The γ rays are extremely penetrating, being capable of passing through large thicknesses of solid matter. For instance, the γ rays given out by very strong radium bromide can be detected after passing through 30 cms. of iron. They are very much more penetrating than the X rays from a very hard X ray bulb. They ionize gases but to a very much less extent than either the α or β rays, the ionization being very approximately proportional to the density of the gas.

No one has as yet succeeded in deviating the γ rays by either an electric or magnetic field. Their great penetrating power and their non-deviability show a close resemblance to very hard X rays. We know also that X rays are produced by the sudden stopping of a moving electron, and it is reasonable to suppose that they would be produced by the sudden starting of an electron. Now experiment has shown that γ rays occur only in conjunction with β rays and the β rays we know are electrons. Consequently, it is reasonable to suppose that the γ rays are, like X rays, electromagnetic pulses produced by the sudden emission of the β particle, or electron from the radio-active substance. This theory seems to be supported strongly by the evidence at present available, although it is very difficult to settle the question definitely by direct proof.

576. Production of Uranium X and Thorium X.—Crookes in 1900 showed that by a simple chemical process he could separate from uranium a constituent which was many times more active photographically than the uranium from which it was separated and, in addition, the separation of this constituent left the uranium photographically inactive. This new and unknown constituent he called Uranium X, or Ur. X. Becquerel obtained similar results using a slightly different chemical process, and, on testing about a year later the Ur. X and the uranium from which it had been separated, discovered in addition the curious fact

that *the uranium had completely recovered its usual amount of activity* while the Ur. X had entirely lost its activity. Rutherford and Soddy later succeeded in performing a similar chemical operation with thorium, separating a very active constituent from thorium, which they called Thorium X or Th. X and which acted in a manner very similar to Ur. X.

These phenomena have been thoroughly examined, both by the photographic and electrical methods, and it has been found in the case of uranium that after separation the Ur. X was very active photographically but inactive electrically, because it gave out β rays but no α rays, while the uranium from which it had been separated was inactive photographically but still active electrically, owing to the fact that it gave out α rays but practically no β rays. The Ur. X gradually lost its activity, while the uranium regained its β ray activity again, and the loss in the one instance and the recovery in the other took place at the same rate. When the Ur. X had lost half its activity the uranium had regained half its original activity and each process took about 20.7 days. The way in which this occurred is shown very clearly by the curves in Fig. 450 which represent the activity of each at different times after separation, the ordinates representing activity and the abscissæ time in days. Similar results but of a slightly more complicated nature have been observed for thorium, but the time taken for the activity of Th. X to decay to half its maximum value and that of the thorium to regain half its activity have been found to be only 6.64 days.

FIG. 450.—(After Rutherford, Radio-activity.)

These results indicate that some process must be continually going on in these substances. Since the Ur. X which gives out β rays can be separated from the normal uranium leaving it devoid of β rays, therefore, the β rays must arise from the Ur. X, and since the uranium regains the β ray activity after separation more Ur. X must be formed in the uranium compound to give rise to these rays. This can be shown to be true, for Ur. X can

be separated a second time after recovery has taken place. The activity of normal uranium does not change, consequently, there must be a state of equilibrium in the uranium in which Ur. X is being formed at the same rate as it dies away, so that the resultant activity remains constant. This is borne out by the fact that the rate of decay of Ur. X is the same as the rate of recovery of the uranium from which it was separated. Processes of a similar nature have been shown to be continually taking place in radium and actinium compounds.

These facts, along with a great deal of additional evidence, some of which we shall consider later, led Rutherford and Soddy to formulate *the theory of successive changes in radio-active substances.* According to this theory the different radio-active substances are gradually undergoing a process of transformation by which they are changing in regular succession from one product to another without the help of any outside agency. We shall see later that Th. X, for instance, is not lost when its activity completely decays, but it disappears as Th. X and changes into another product or substance. Most of these *transformation products* as they are called give out radiations similar to those we have considered, but some do not give out any at all and are, consequently, called rayless products. The rates at which these changes take place vary very greatly for the different products, some changes only taking a few seconds to complete, while others extend over hundreds of years. The time it takes any one of these changes to be *half* completed is generally spoken of as the *period* of that transformation, as this time is usually much more easily determined experimentally with accuracy than the time of the complete change.

Actinium possesses a corresponding active constituent called Actinium X with properties similar to Th. X.

577. Emanations from Radio-active Bodies.—The early experimenters on thorium observed that the radiations given out by thorium compounds were very irregular. Rutherford investigated this irregularity and found that it was due to the emission of some sort of radio-active particles from the thorium compound. To these particles he gave the name "emanation," and he found that it was not like the radiations which we have already considered, but *acted in all respects like a gas.* It will diffuse through porous solids and through gases and it may be carried away by a current of air. It is capable of ionizing a gas itself and of acting on a photographic plate.

It does not itself consist of ions but has the power of producing ions in the gas, for it may be passed through cotton wool or bubbled through solutions without losing its power of ionising a gas. This differs from a gas ionised in the ordinary way, for the gas will lose its ions under these circumstances while the emanation does not.

The emanation is not affected by an electric field. The electric field removes the ions produced by it but does not remove the emanation itself. The emanation cannot, therefore, consist of charged particles like the ions.

Both radium and actinium compounds give out an emanation possessing properties similar to the thorium emanation, but as far as is known at present uranium compounds do not give off any emanation.

These emanations are chemically inactive, not being affected by the strongest reagents. They are not altered by being passed through a platinum tube raised to a white heat, nor by being cooled to the temperature of solid carbon dioxide. The emanations can be condensed when passed through a tube immersed in liquid air. This is a very important and crucial experiment, proving conclusively the gaseous nature of the emanation.

Actinium emanation may be condensed under the same conditions as thorium emanations.

If the emanation be removed from the thorium, by drawing off into another vessel both it and the air with which it is mixed, its activity dies away very rapidly with time. Also if a quantity of thorium be placed in a closed vessel and the ionisation current measured immediately and at short intervals it is found to gradually rise and finally reach a steady state. The rate at which the current rises in the closed vessel is exactly the same as the rate at which the separated emanation dies away. We have here a state of things similar to the case of thorium and Th. X where the activity of one rises at the same rate as the other dies away. An equilibrum state is reached when the emanation is produced as fast as it dies away.

The emanation is not produced directly by the thorium but is a product of Thorium X. Rutherford and Soddy have shown that when the Th. X is separated from the thorium the latter does not give off any emanation but gradually regains its emanating power. The separated Th. X, however, possesses strong emanating power but gradually loses it. These processes take place at exactly the same rate as the loss and regain of activity by the Th. X and thorium respectively, which we have already considered. This accounts for the decay of the Th. X as it is continually changing into emanation. The emanation and Th. X are distinct substances having distinct properties. These emanations resemble somewhat the rare gases found in the atmosphere, being very inert chemically. Radium emanation is now definitely recognised as an element having an atomic weight of 220 and it has been given the name niton.

578. Excited Activity.—If a solid body be exposed in a closed vessel to the emanations from radium, thorium or actinium its surface becomes coated with an extremely thin solid deposit of very radio-active material. This active deposit is invisible, even under a microscope, but can be dissolved by certain acids and when the solvent is evaporated again it is left behind. It

emits radiations which affect a photographic plate and ionise a gas. If a negatively charged wire be placed in a closed vessel containing the emanation the active deposit is all concentrated on this wire instead of being distributed on the interior of the vessel. By this means a very small wire may be made intensely radio-active.

The active deposit can be removed from a wire by rubbing with sand paper, but the quantity deposited is so extremely small that no increase in weight can be detected in a wire which has received an active deposit. This active deposit is not due to any action of the radiations given out by the radio-active compound but is a direct result of the presence of the emanation, for when no emanation is present no active deposit is observed and, in addition, the amount of excited activity is always proportional to the amount of emanation present.

If the negatively charged wire be exposed to the emanation for several hours and then removed and its activity tested at intervals, it is found to gradually die away with time, according to a law exactly similar to that for the decay of the emanations. The excited activity from thorium decays to half value in about 10.6 hours. It requires time for the excited activity to be deposited on the wire and the deposit increases until it reaches a maximum. This rate of increase is the same as the rate of decrease of activity when the wire is removed from the emanation. There must, consequently, be an exactly similar process going on here as we observed in connection with thorium and Th. X, and with Th. X and the emanation. Just as Th. X is continuously changing into the emanation the emanation is gradually changing into the active deposit and this in turn must be changing into something else.

If the wire be exposed to the thorium emanation for only a few minutes instead of several hours a different phenomenon is observed after removal of the wire. Instead of beginning to decay immediately after removal the activity, which at first is very small, gradually increases until it reaches a maximum in about four hours, and then it decays again at just the same rate as the activity for a long exposure decayed. When the exposure is a long one no initial increase is observed. Rutherford suggested that the active deposit, instead of being one substance, is really made up of two distinct substances one of which is changing into the other. He called these two substances thorium A and thorium B and supposed that thorium A arose from the emanation and was deposited on the wire and then changed into thorium B, and then the thorium B changed into something else. For a short exposure the deposit will consist almost entirely of thorium A, as very little has had time to change into thorium B, and if we suppose that thorium A either gives out no rays at all or rays which produce a very small amount of ionization compared with those from thorium B, then the activity at first will be very small, due almost entirely to the very small portion of thorium B. As thorium A changes into thorium B the activity will increase until the change of A into B just balances the decay of B. Then, as more atoms of B will change per second than are produced from A, the activity will gradually decay. In the case of the long exposure this maximum has been

reached before the wire is removed and tested and, consequently, the initial rise is not observed. Recent investigations show that this active deposit is more complex than was at first supposed. It is now known that it consists of several distinct substances.

An examination of the active deposit from radium shows that the transformations taking place are more complicated than those of thorium and actinium. The decay curves when measured by the α rays are quite different from those obtained by the β or γ rays. The two latter give identical curves showing that the β and γ rays occur together. By a process of analysis similar to that used for thorium it has been shown that the active deposit from the radium emanation consists of even a greater number of distinct transformation products than is produced by thorium or actinium.

579. Heat Emitted by Radium and Thorium.—Curie and Laborde discovered that radium is always hotter than its surroundings and emits heat at the rate of 100 calories per gram per hour. It has also been found that thorium acts similarly though in a minor degree. This is readily explained by the high velocity and kinetic energy of the α particles (§572) and the readiness with which they are absorbed (§574). Many of the particles that start within the radio-active body are absorbed by the body itself and their kinetic energy is transformed into heat.

580. Theory of Radio-active Changes.—We have seen that in the radio-active bodies continuous changes from one substance to another are taking place which so far have never been observed in any other class of materials. Each of these substances is entirely distinct from the others and has distinct physical and chemical properties. They, however, gradually decay and each one has a distinct and definite period of decay which distinguishes it from all the others. How do these changes come about? The disintegration theory or theory of successive changes furnishes the now generally accepted explanation.

According to the theory of J. J. Thomson atoms may be considered complex structures consisting of systems of positively and negatively charged particles in very rapid rotation and held together by their mutual forces in equilibrium. According to the disintegration theory this complex structure constituting the atom of radium (which we shall take as a typical example) becomes by some means unstable and one of the positively charged α particles is suddenly expelled with great velocity. The structure of the atom which remains is now different and constitutes the atom of a new substance, namely, the emanation.

The atoms of the emanation are unstable and gradually change by the expulsion of another α particle, leaving a new structure, namely, the atom of radium A, and the process is continued throughout the successive changes. The processes are not identical in all instances, for in some cases an α particle alone is expelled, but in others β particles are expelled accompanied by γ rays, while in others all three types are given out.

Why do these atoms suddenly become unstable and break up without any apparent cause? Several explanations have been offered to account for this, but the most probable one seems to be that if this system of charged particles, of which the atom probably consists, is in rapid rotation it must be radiating energy, and when sufficient energy has been radiated the mutual forces of the system no longer balance and one or more of the particles escape and cause disintegration. These atoms have an independent existence and distinct physical and chemical properties, but they differ from the atoms of ordinary non-radio-active elements in the fact that they are not permanent. To distinguish them from ordinary atoms the term metabolon has been suggested as a suitable name.

A few of these transformation products do not emit any rays at all and the change from them into the succeeding substance apparently takes place without the expulsion of any particles. These so-called rayless changes may be explained in either of two ways. The new product may be formed in this case simply by a rearrangement of the system of charged particles, but not with sufficient violence to expel any of the system, or it may be produced by the expulsion of one or more particles, but with a velocity too slow to ionize the gas. It has been shown that when the velocity of the α particle falls below 10^9 cms. per second it ceases to ionize the gas, and consequently an α particle expelled with a velocity below this minimum would escape detection since no ions would be produced.

This latter hypothesis suggests that all matter may possibly be undergoing a slow change in a similar manner, and that the reason this change has been observed only in the so-called radio-active bodies and not in other non-radio-active bodies is that in the case of the radio-active bodies the charged particles are expelled with sufficient violence to ionize the gas while in other bodies they may be expelled but not with sufficient velocity to produce ions.

581. Radio-active Elements.—The following table contains a summary of all the active products at present known. On account of the incomplete state of the subject this list will in all probability undergo in the future slight changes as a result of further investigation.

TABLE OF RADIO-ACTIVE ELEMENTS

Radio-active products	Transformation period	Nature of rays emitted	Radio-active products	Transformation period	Nature of roys emitted
Uranium I ↓	5×10^9 years	α	Thorium ↓	1.8×10^{10} years	α
↘ Uranium Y	1.5 days	β	Mesothorium I ↓	5.5 years	No rays
Uranium X_1 ↓	24.6 days	β	Mesothorium II ↓	6.2 hours	β
Uranium X_2 ↓	1.15 min.	β	Radiothorium ↓	2.02 years	α
Uranium 2 ↓	2×10^6 years	α	Thorium X ↓	3.64 days	α
Ionium ↓	2×10^5 years[1]	α	Thorium emanation ↓	54 sec.	α
Radium ↓	1730 years	α, β	Thorium A ↓	.14 sec.	α
Radium emanation (Niton) ↓	3.85 days	α	Thorium B ↓	10.6 hours	β
Radium A ↓	3 min.	α	Thorium C_1 ↓	60 min.	α, β
Radium B ↓	26.7 min.	β	↘ Thorium D	3.1 min.	β
Radium C_1 ↓	19.5 min.	α, β	Thorium C_2	10^{-11} sec.	α
↘ Radium C_2	1.4 min.	β	Actinium ↓	....	No rays
Radium C′ ↓	10^{-6} sec.	α	Radio-actinium ↓	19.5 days	α
Radium D ↓	16.5 years	β	Actinium X ↓	11.4 days	α, β
Radium E ↓	5 days	β	Actinium emanation ↓	3.9 sec.	α
Radium F (Polonium)	136 hours	α	Actinium A ↓	.002 sec.	α
			Actinium B ↓	36.1 min.	β
			Actinium C ↓	2.15 min.	α
			Actinium D	4.71 min.	β, γ

[1] Very recently another product called Ur Y has been discovered. It has a period of 1.5 days and emits soft β rays and probably α rays. It is considered to be a lateral disintegration product of uranium.

References.

THOMSON'S *Conduction of Electricity through Gases.*

STRUTT'S *The Becquerel Rays and Other Properties of Radium.*

RUTHERFORD'S *Radio-activity.*

RUTHERFORD'S *Radio-active Transformations.*

MCCLUNG'S *Conduction of Electricity through Gases and Radio-activity.*

SOUND

By A. Wilmer Duff, D. Sc.

Professor of Physics in the Worcester Polytechnic Institute, Worcester, Mass.

NATURE AND PROPAGATION OF SOUND

582. Sources of Sound.—On hearing a sound we instinctively think of its origin and we are usually able to trace it to some body which we call the source of the sound. To ascertain how a body produces a sound we may take a body that can be kept sounding while it is being observed. If a large bell be made to produce sound by striking it, or a glass jar by stroking it with a violin bow, a light pendulum hung against the bell or jar will be kicked away at each contact. A metal rod clamped at the middle and sounded by rubbing one-half of it with a rosined cloth or glove will give violent blows to a pendulum hung in contact with the other end. If a violin be held in the hand and one of the strings be plucked or stroked by a bow, the hand will tell us that the wooden body of the violin is vibrating, and, while the vibrations of the body cannot be seen, those of the string are clearly visible.

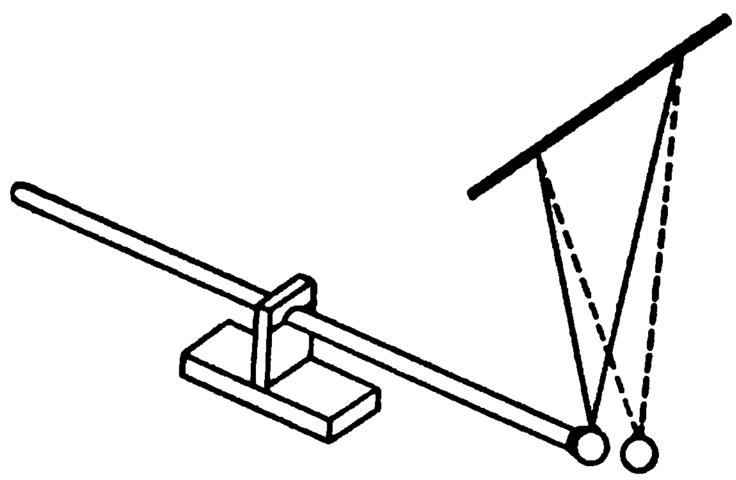


Fig. 451.

In all cases in which the action can be observed closely enough, sounding bodies are found to be in a state of *vibration*. In many cases the vibrations are so small or cease so quickly that they cannot be detected. But there is no doubt that the head of a hammer vibrates for a moment when it strikes a nail and the board into which the nail is being driven also vibrates for a moment.

583. Media in which Sound Travels.—It is well known that the clearness with which distant sounds are heard depends on the

state of the atmosphere and the direction of the wind. Hence the air is the ordinary medium of transmission. Sound will not travel in a vacuum, as can be readily shown by placing an electric bell under the receiver of an air pump. The sound will diminish as the air is removed, but it will be restored if the air or any other gas is allowed to enter.

Sound is also transmitted by liquids and solids. If two stones be struck together under water a loud sound will be heard by an ear held beneath the surface. A watch placed on one end of a long table can be heard by an ear pressed against the other end of the table. Miners imprisoned by an accident in a mine sometimes send signals to the outer world by blows of their picks on the rock. The approach of a distant train or the galloping of horses can be heard by an ear pressed against the ground. Beethoven, who was deaf, heard some of his own compositions only by means of a stick, one end of which rested against the sounding board of the piano while the other was pressed against his teeth.

584. The Nature of Sound.—Since the sources of sound are vibrating bodies, it follows that sound travelling through a medium must be a wave motion. That it does not consist, like odors, of particles transmitted from the source of sound is shown by the fact that sounds of great intensity can come to us through a layer of smoky air without bringing the smoke with it.

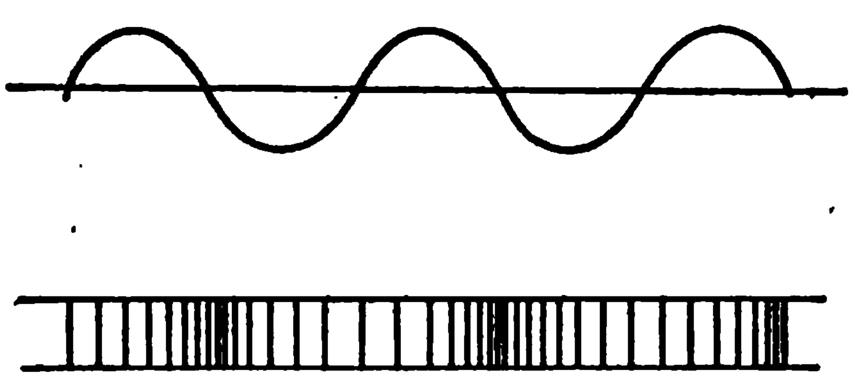


Fig. 452.

Wave motions may be either transverse, like the to and fro motions of a cord, or longitudinal, like the compressions and extensions of a spiral spring, or they may be a mixture of both, as in the case of water waves. Now a gas offers elastic resistance to compression and can, therefore, transmit longitudinal waves; but it offers no sustained resistance to changes of shape, or shears, and it cannot, therefore, transmit transverse elastic waves. Hence sound waves are longitudinal waves or waves of compression and dilatation.

To represent sound waves graphically we make use of the

device explained in §238, that is, we take the axis of abscissæ in the direction in which the waves are travelling and draw ordinates to represent the displacements of the air particles at the corresponding points, although the displacements are really in the direction of propagation. These actual displacements are in reality excessively minute, ranging from 2 or 3 mm. in the loudest sounds to about 10^{-7} mm. in sounds that are just audible. Hence the ordinates are drawn on a greatly enlarged scale. The forward motion of the wave is represented to the mind by supposing the curve to move forward with the velocity of the wave. The curve crosses the axis of abscissæ at the middle of a condensation or of a rarefaction and the maximum ordinate occurs at a place of no condensation or rarefaction. Sound waves are sometimes represented by similar curves the ordinates of which represent the degrees of condensation (positive or negative) at points in the wave. Such curves are half a wave length ahead of (or behind) those that represent displacements.

That sound has all the properties characteristic of wave motions in general will appear in the paragraphs which immediately follow.

585. Velocity of Sound.—When a man is seen chopping a log or hammering a nail at a great distance, the sound of each blow is not heard until some time after the blow is seen. Steam may be seen to issue from a distant whistle before the sound is heard. Fire-alarm whistles sounded simultaneously may be heard separately if the observer is not at equal distances from the stations. Lightning precedes thunder. From these facts it is evident that sound travels at a definite rate which can be measured. That this rate is not as great as the velocity of shells fired by modern high power guns is shown by the experiences of soldiers in the trenches in the European war. The sound made by a passing shell was heard before that of the firing of the gun arrived.

The velocity of sound has been determined by methods suggested by some of the above experiences. The most accurate of such direct determinations were made near Utrecht in 1823 and near Paris in 1822, and gave a mean value of about 341 meters per second at 16°C. The sounds observed were those of the firing of cannon at great distances and care was taken to reduce the effect of wind by alternate observations of sounds travelling

in opposite directions. The velocity of sound in water was found in 1827 by means of bells sounded beneath the water of Lake Geneva in Switzerland. But we shall see later that there are more readily available laboratory methods for finding the velocity of sound in gases, liquids, or solids.

That sounds of different pitch travel at practically the same rate is shown by the fact that music made by a band can be heard as music at a considerable distance; for if notes sounded at the same time were not heard simultaneously both harmony and melody would be distorted. There is some evidence that very loud sounds may travel at somewhat greater velocity than others, but the difference, if it exists, is slight.

586. Formula for the Velocity of Sound.—From the principles of Mechanics applied to waves of compressions and rarefactions Newton derived the formula which has been stated in §251, namely,

$$v = \sqrt{\frac{E}{\rho}}$$

ρ being the density of the medium and E its elasticity. Now the elasticity of a gas at constant temperature, or its *isothermal* elasticity, is equal to the pressure, p (§223). But if we substitute p, expressed in absolute units (dynes per square centimeter) in the above we get a result which is much too small, as Newton found. This difficulty was not removed until Laplace pointed out that the temperature must be momentarily elevated in a compression and depressed in a rarefaction as a train of sound waves passes, and that E should therefore be taken as the *adiabatic* elasticity, which, as we have seen (§346) is κp, where κ is the ratio of the specific heats of the gas at constant pressure and at constant volume respectively. With this condition the formula becomes

$$v = \sqrt{\frac{\kappa p}{\rho}}$$

and this is found to agree with experimental results.

From this formula it is evident that the velocity of sound in a gas is independent of the pressure or density, provided the temperature is constant, since, in accordance with Boyle's law, p is proportional to ρ. Thus at the same temperature the velocity is the same at the top of a mountain as in a valley. In gases of

different densities the velocities are inversely as the square roots of the densities. Thus the velocity in hydrogen is four times that in oxygen. Since the density of water vapor is less than that of air, the presence of water vapor in air, at given atmospheric pressure, causes a slight decrease in the velocity of sound.

From the above formula we can also find how the velocity of sound depends on the temperature of a gas. For from the general gas law (§280), since density varies inversely as volume, we see that $p/\rho = (p_o/\rho_o)(1+at)$. Hence

$$v = \sqrt{\kappa\frac{p_o}{\rho_o}(1+at)} = v_o\sqrt{(1+at)}$$

where $v_o = \sqrt{\kappa\frac{p_o}{\rho_o}}$ is the velocity at 0°C.

VELOCITY OF SOUND IN DIFFERENT SUBSTANCES (IN METERS PER SECOND) AT 0°C.

Air	332	Steel	4975
Hydrogen	1268	Lead	1420
Carbon dioxide	261	Glass	4860
Fresh water	1435	Pine wood	3300
Sea water	1454	Walnut wood	4800
Mercury	1484	India rubber	5000

587. Mechanical Effects of Sound Waves.—Sound waves produce mechanical effects by which they can be detected and to some extent measured. A very simple and useful detector is a "sensitive flame," that is, a long slender gas jet issuing from a very fine nozzle (a glass tube drawn out to a fine point) under steady pressure. If the flow is regulated until the flame is just about to flare or become unsteady, sounds of high pitch falling on the orifice will cause instability and the flame will shorten greatly and "roar" like an ordinary gas flame when the stopcock is too wide open. A hiss or the rattling of a bunch of keys is especially effective. A Welsbach flame turned very low can be made very sensitive to high pitched sounds (like those on the highest string of a violin).

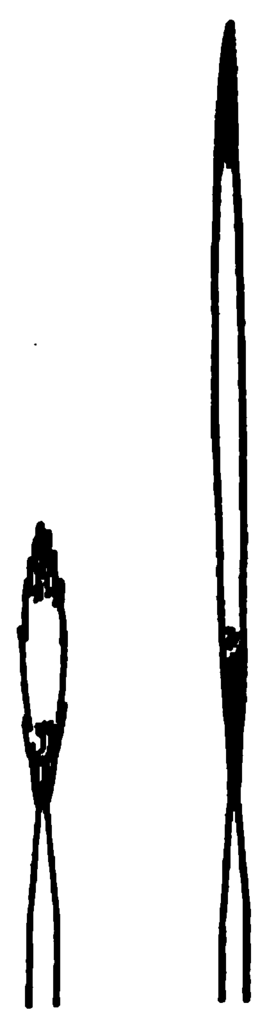


FIG. 453.

A so-called "manometric flame" is a gas jet fed by gas which passes through a small box one side

of which is of thin rubber. Sound waves falling on the rubber produce variations of pressure in the gas and corresponding variations in the height of the flame. If the latter be viewed in a rotating mirror, a band of curved outline corresponding to the peculiarities of the sound will be seen.

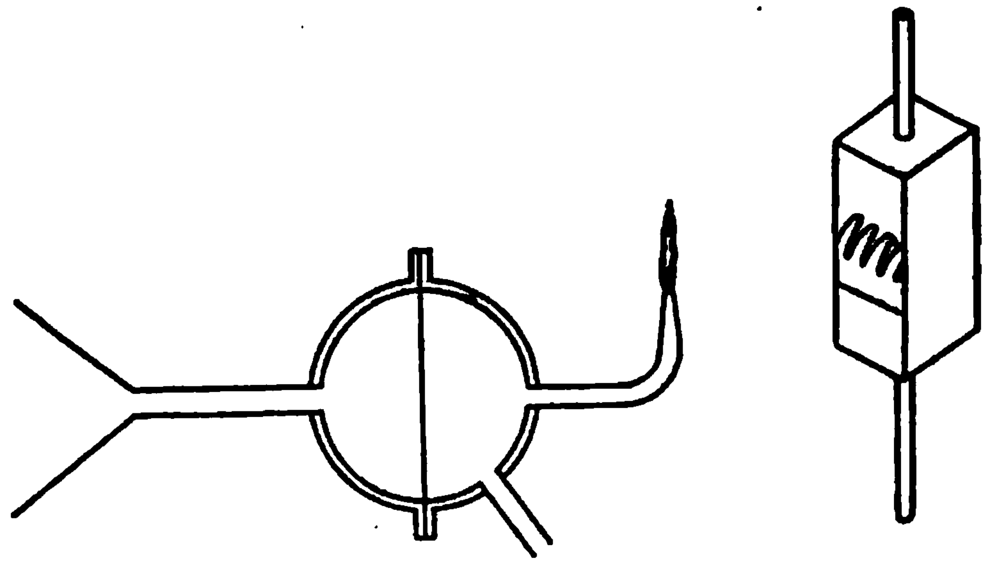


FIG. 454.

The vibrations of a disk on which sound waves fall can be made visible by reflecting a beam of light from a small mirror connected with the disk. This method has been very successfully used by Professor D. C. Miller in his "phonodeik," a fine fiber *f* attached at one end to a thin glass plate *p* is wrapped around a tiny spindle which carries the mirror *m*. A record obtained by Miller will be given later (Fig. 465).

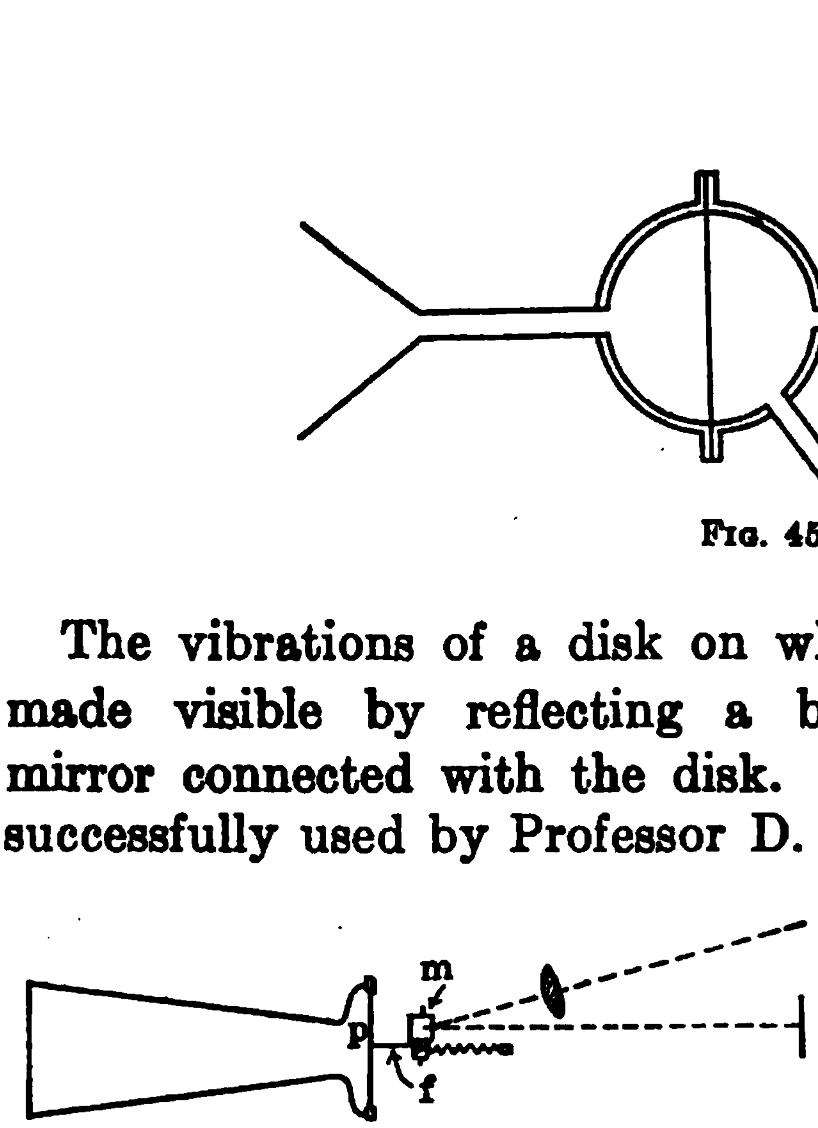


FIG. 455

The effect of sound waves on the ear consists of a vibration of the ear drum produced by the alternating pressure of the waves. Their effect on the disk of a telephone is similar. The vibrations of the disk produce alternating electric currents in the line (§525) and these can be shown by a vibration galvanometer (§438) or oscillograph (§524), or they can be rectified by a crystal rectifier (§545) and shown by an ordinary galvanometer.

Sustained waves of any kind carry a steady stream of energy from the source and falling on a surface produce a sustained pressure. In the case of sound waves this pressure is small. It can be shown by concentrating the loud sound produced by a stream of sparks from an induction coil by means of a large metallic mirror. A small vane similar to that of a radiometer placed at the focus will be set in rotation. Waves falling on a small disk in a resonating tube (Fig. 456) tend to set it at right angles to the stream by an action somewhat similar to that described in §204. This action has been used in

certain investigations on the energy flow in sound waves, by Lord Rayleigh and others. No instrument of general usefulness in the measurement of sound intensity, that is, of the energy flow, has yet been devised.

588. Photographs of Sound Waves.—Sound waves can be photographed by a method very analogous to the method used

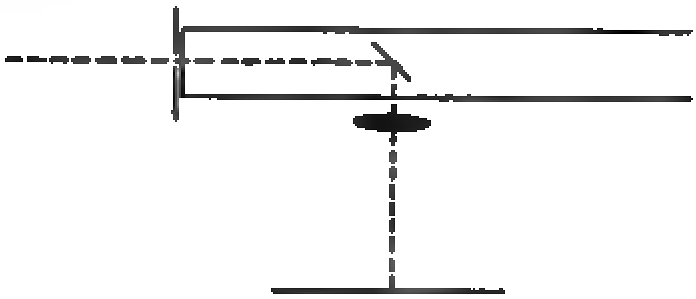


FIG. 456.

in projecting ripples (§256). The waves pass in front of a photographic plate in the dark; a single spark from a distant Leyden jar illuminates the plate for a moment and leaves a record of the condensations and rarefactions owing to the refraction of the light as it passes through the sound wave. Fig. 457 is such a

FIG. 457.

record (by Foley and Souder) showing the reflection of a spherical wave of sound by a plane surface. (See also §691).

589. Reflection of Sound.—Sound, like other wave motions, is reflected when it falls on a suitable surface, and in the process it follows the same laws as light and radiant heat, namely, the reflected and incident rays lie on opposite sides of the normal to

the surface and make equal angles with it. The identity of these laws for sound and light can be shown by means of two large concave metallic mirrors placed opposite each other with their axes in the same line. In front of each mirror there is a point,

FIG. 458.

its principal focus, such that a bright light placed at one focus produces a bright image at the other. When these points have been found, a source of sound of high pitch placed at one focus will produce an intense sound at the other focus, as may be seen by its action on a sensitive flame. If a singing arc light be placed at one focus, the light, heat, and sound will be focussed simultaneously by the other mirror, as may be ascertained by means of a small screen, a thermopile (§333), and a sensitive flame respectively.

Reflection of sound gives rise to echoes. The echo from a large reflecting surface, such as the side of an isolated house, is heard most distinctly when the observer is in a line perpendicular to the surface, and it decreases rapidly as he moves away from that line. Most of the sound heard in an auditorium is reflected sound. This will be referred to more fully later.

590. Refraction of Sound.—Waves are refracted, or their line of propagation is bent when they pass obliquely from one medium into another in which the velocity is different (§255). The refraction of sound cannot be shown satisfactorily on a small scale by lecture or laboratory apparatus, but it takes place on a large scale in nature. The chief causes of such refraction are winds and variations of temperature in the atmosphere.

The velocity of a wind is less nearer the surface of the earth than higher up, since near the surface it is retarded by the frictional resistance of the surface. When sound is travelling in the same direction as the wind its resultant velocity is greater above

than below. Hence the waves, which always travel at right angles to their fronts, are tilted forward, or the direction of their motion is deflected downward (Fig. 459, *B*). The opposite effect takes place when the sound is travelling against the wind (Fig. 459, *A*). This explains why sound is better heard with the wind than against it and why it is an advantage in the latter case to listen from an elevation.

Similar effects result from the temperature being different at different heights in the atmosphere. Usually on fine days the temperature is lower at an elevation and the velocity of sound is less there. Thus the waves are deflected upward as in Fig. 459, *A* and hearing near the surface is poor. At night or near sunrise or sunset the surface is colder and the gradient of temperature is less and sounds are heard better. Moreover on a hot day the air is heated irregularly by contact with parts of the surface at

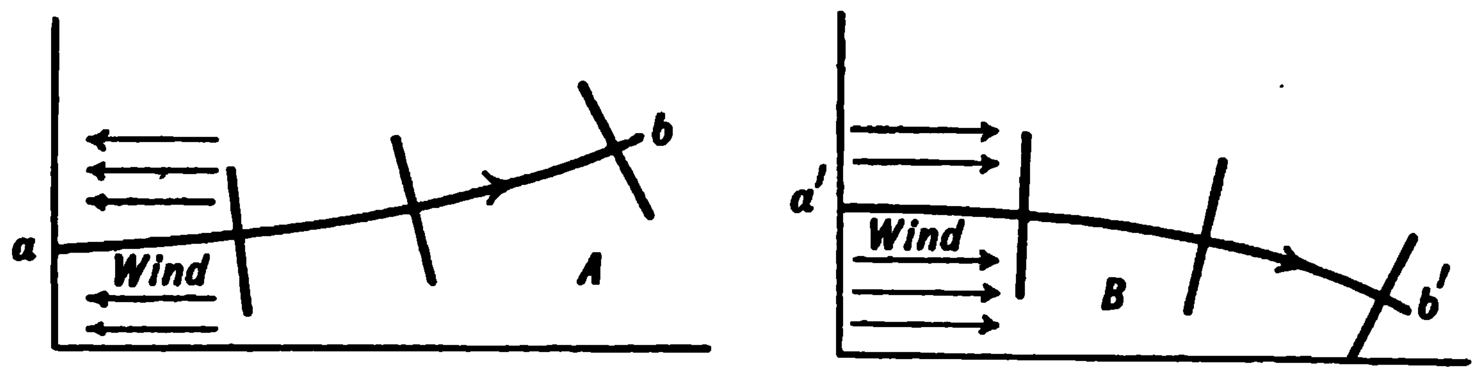


FIG. 459.

different temperatures and there are numerous irregular convection currents in the atmosphere. These cause irregular reflections and refractions of sound as in the similar case of light passing over a heated stove. These effects also help us to understand why distant sounds are frequently heard most clearly before a storm.

591. Diffraction of Sound.—Diffraction means the bending of waves around obstacles (§256). It prevents the formation of sharp shadows. The amount of diffraction in any case depends on the linear dimensions of the obstacle compared with the wave length of the sound. A hill casts a fairly definite sound shadow because it is large in comparison with the wave length. Cases have been known in which houses in the shadow of a hill have suffered no damage from very loud sounds, such as the explosion of a powder magazine or the firing of cannon, while the windows of equally distant houses not in the shadow were broken by the

impact of the sound waves. But small obstacles, such as trees and posts, cast no observable sound shadows, except when tested by wave of very short length. The human head is of sufficient size to cast something of a shadow for sounds of high pitch. If the sound is to the left or the right of the observer, one ear is in this partial shadow and from the relative intensities as heard by the two ears we judge of the direction of the source. Long waves produce about equal intensities at the two ears, but at one ear the phase of the waves is later than at the other and it has been shown that it is by this slight difference that the mind unconsciously judges of the direction of the source of the sound.

592. The Phonograph and the Grammophone.—When we speak against the middle of a thin elastic disk such as is used in the telephone transmitter, the disk vibrates backward and forward in unison with the sound waves. In the phonograph these vibrations are recorded on a drum of hard wax that is kept in rotation behind the vibrating disk. A short needle attached to the vibrating disk presses on the moving drum and the vibrations cause it to plough a fine furrow in the wax. Thus the furrow is a record of the vibrations of the disk. The sound can be reproduced by allowing the needle to travel again along the furrow. It is thus pushed up and down, following very closely the motion that produced the furrow, so that it causes the disk to repeat its original vibrations and the disk therefore reproduces the original sound. (It is found better to use different disks and needles for recording and reproducing.)

The principle of the grammophone is essentially the same, but the needle is connected to the center of the vibrating disk by a lever in such a way that it moves sidewise and not up and down. It thus produces a transverse furrow on a rotating disk.

MUSICAL SOUNDS

593. Characteristics of Musical Sounds.—The ear is remarkably acute in distinguishing minute differences in sounds. In doing so it takes note of three fundamental properties in which sounds differ, namely, *loudness, pitch,* and *quality*. These we may call the three characteristics of musical sounds. They and the direction from which the sound comes are all that the ear can

tell us about it. From these the mind can, as a result of long experience, draw very rapid conclusions as to the nature of the source. The words loudness, pitch, and quality need no definition, as we are more familiar with them than with the terms which might be used in defining them. But we must remember taht the words stand for sensations. Sound waves must have corresponding characteristics that account for these differences in the sensations of sound.

594. Loudness.—If we strike a bell, a drum, or a violin string very gently, the vibrations of the instrument will be of small amplitude and the sound will be weak, but a stronger blow will produce vibrations of greater amplitude and the sound will be louder. Now it is evident that the amplitude of the air vibrations is greater the greater the vibrations of the source. Hence the loudness of the sound heard depends on the amplitude of the vibrations in the waves. The same conclusion is reached by considering that the sound heard is weaker the farther the hearer is from the source. The energy that falls on the ear must be less at greater distances and therefore the amplitude of vibration must be less.

The intensity of sound waves means, as in the case of waves of any kind (§259) the flow of energy per unit time per unit area perpendicular to the direction in which the waves are travelling. For simple harmonic waves of a given frequency the intensity is proportional to the square of the amplitude (§§112, 259); it is also proportional to the square of the frequency. Hence in the case of strictly spherical waves spreading from a point source the intensity would vary inversely as the square of the distance provided the waves were not damped (§259). But, owing to the presence of reflecting surfaces, sound waves produced under ordinary circumstances do not remain spherical. There is also some damping, due to viscosity of the medium, which may have considerable effect for great distances of transmission out of doors, and, as already noted, lack of homogeneity of the atmosphere causes dissipation. For these reasons the inverse square law is not applicable to sound waves under ordinary conditions of hearing.

The intensity of sound waves and the loudness of the sensation they produce are related, but the relation is not a simple one.

The loudness of the sensation also depends on the pitch of the sound. But such questions belong to Psychology rather than Physics.

595. Pitch.—The sound of a toy-whistle is of high pitch or shrill, that of an automobile horn is of medium pitch, while the tones of a church bell are deep or of low pitch. The physical cause of these differences of pitch may be shown by comparing the different sounds that can be produced by drawing a card along the teeth of a comb. If it be drawn slowly a sound of low pitch will be heard, but if it be drawn as rapidly as possible the sound will be of high pitch. Every time the card slips off one tooth and strikes the next an impulse is given to the air. The more numerous these impulses per second, that is, the more air-waves are started per second, the higher the pitch of the sound will be. Similar results are produced when a machine saw cuts a board. The rise and fall of the pitch of the sound is due to the variations

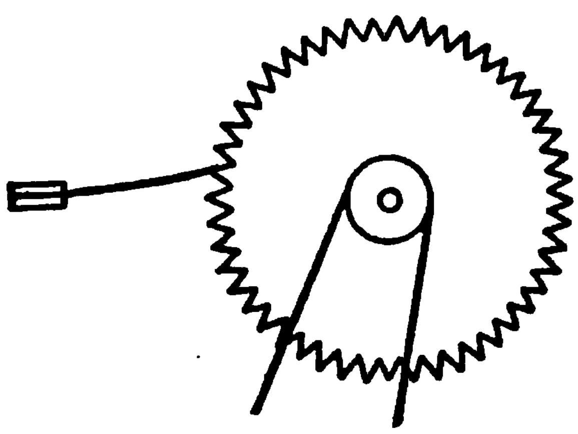


Fig. 460.—Savart's wheel.

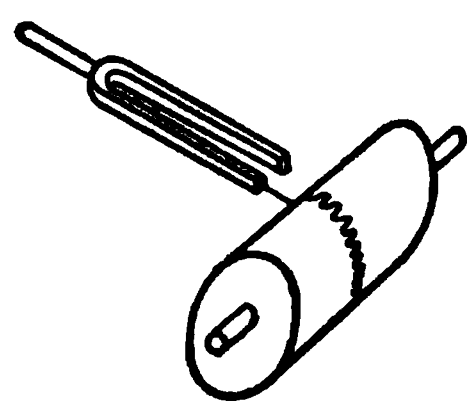


Fig. 461.—Tuning-fork record.

of the speed of the saw. Savart's toothed wheel for showing the cause of the pitch of sounds illustrates the same principle.

If a phonograph be driven at lower than its normal speed the pitch of every note will be lowered in correspondence with the decrease of frequency of the impulses imparted to the membrane.

A tuning-fork with a small spring stylus attached to a prong can be made to inscribe its vibrations on a revolving lamp-blacked drum or on a lamp-blacked sheet of glass drawn beneath the fork. It is found that for a certain speed of drum or glass the number of waves recorded is greater the higher the pitch of the fork. When this experiment is made with sufficient care the frequency of the fork can be accurately measured.

From the above we conclude that *the pitch of a sound depends on the frequency of the vibrations in the sound wave.*

Now we have already seen that there is a simple relation between frequency and wave length in a medium; they vary in inverse ratio, that is, the longer the waves the less the frequency. Hence we may also say that *the pitch of a sound depends on the length of the sound wave.*

It must not be concluded from the above that a regular succession of air-waves produces the sensation of sound, no matter what the frequency. The enormously rapid vibrations of the wings of a mosquito produce a sound of very high pitch and the slower vibrations of a humming-bird's wings produce a sound of low pitch; but no note of definite pitch is produced by the flight of a swallow. In fact, to produce sounds of definite pitch the frequency of the waves must not be less than about 20 per second. On the other hand when their frequency exceeds about 20,000 per second air-waves do not produce the sensation of sound at all, though their existence may be shown by their action on a sensitive flame and other means. The upper limit of frequency is different for different persons and is lowered by advancing age and by disease of the ear.

FIG. 462.—Disk syren.

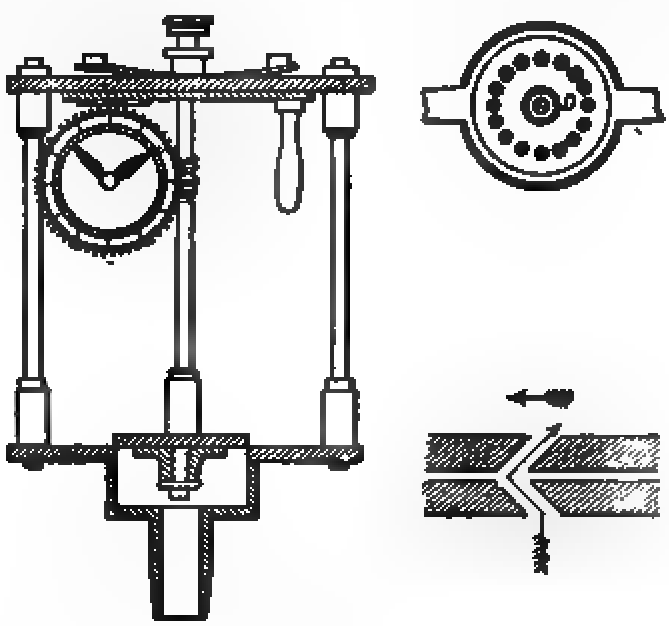


FIG. 463.—Recording syren.

The syren is an instrument for producing sounds of definite pitch by a succession of puffs of air following one another in rapid

succession. In the simplest form of syren a circular disk with circular rings of holes is driven at a high speed and a puff is produced every time a hole comes opposite the end of a tube through which air is driven under pressure. When the frequency of succession of the puffs is great enough to produce a note, the pitch can be raised by increasing the speed of the disk or by transferring the tube to a ring in which there is a larger number of holes. In the more complete form of syren the rotation is produced by the compressed air as it escapes from a box the cover of which is a disk with a ring of holes corresponding to those in the rotating disk. The holes in the rotating disk and those in the fixed disk slope in opposite directions and the jets of air impinge obliquely on the sides of the holes of the rotating disk thus causing it to rotate. The frequency of the note is found by means of a suitable speed counter geared to the rotating disk.

596. Doppler's Principle.—When an observer is *in motion* toward a source of sound, the pitch of the note heard is higher

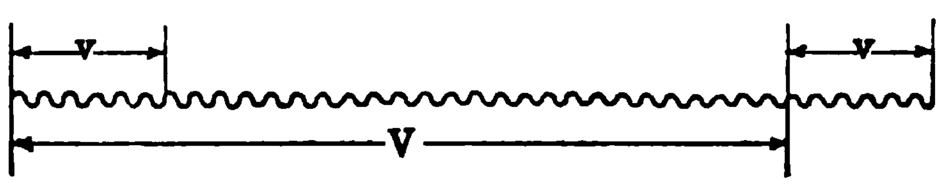


FIG. 464.

than when he is at rest. If the hearer is in motion away from the source, he hears a lower note than when he is at rest. Similar results follow when the source is in motion toward or away from the observer. The pitch of the gong of a fire engine or of the whistle of a locomotive is higher when the source is approaching the hearer than when it has passed and is receding.

When the observer is in motion toward the source, he receives more waves in each second than when he is at rest. The additional waves received are those which occupy the distance, v, which he traverses in a second, and, if λ is the wave length, these are v/λ in number. If V be the velocity of sound and n the frequency of the source, $V=n\lambda$. Hence the increase in frequency heard is $\frac{v}{V}n$ and the pitch as heard is therefore $\left(n1+\frac{v}{V}\right)$. For the case in which the hearer is in motion away from the source the sign of v must be reversed.

When the source is in motion toward the hearer the effect is a shortening of the wave length, for the source is following after the approaching waves, and the crests therefore come closer together. If the frequency of the source is n and its velocity is v, during each vibration it travels a distance v/n and each wave length is shortened by this amount. Hence the wave length of the sound heard is not $\lambda = \frac{V}{n}$ but $\lambda' = \left(\frac{V}{n} - \frac{v}{n}\right)$ and the frequency of the note heard is $\frac{V}{\lambda'} = \frac{V}{V-v}n$. If the source is receding the sign of v must be reversed.

The above two expressions do not differ appreciably if v is small compared with V.

If a vibrating tuning fork on its sounding box be moved rapidly toward a wall or blackboard, an observer will hear two notes of different pitch. One is the note heard directly from the receding tuning fork and is lowered in pitch by the motion. The other note is due to the waves reflected from the wall and this is raised in pitch. The interference of these two notes produces beats.

597. Scale of Musical Sounds.—The difference of pitch of two sounds is called the *interval* between them. In the practice of music these intervals are learned by ear, but in scientific work an interval is stated by giving the ratio of the frequency of the higher sound to that of the lower. Certain intervals have received particular names. Thus the interval between two sounds whose frequencies are as 2 : 1 is called an interval of an *octave*.

A certain succession of notes intermediate between a note and its octave has been found suitable for musical purposes and is called a *musical scale* (major diatonic scale). These notes are called by the letters of the alphabet from A to G. Notes which are each an octave above one of these are called by letters from a to g, those still higher are called a' to g', a'' to g'', etc.

9/8		10/9	16/15	9/8	10/9	9/8	16/15
c	*d*	*e*	*f*	*g*	*a*	*b*	*c*
24 :	27 :	30 :	32 :	36 :	40 :	45 :	48
261	294	316	348	391	435	489	522
256		320		384			512

In the above the second line contains the letters for these notes

(in the scale of C major) and in the first line the fractions that measure the intervals between consecutive notes are given. In the third line is a series of numbers such that the ratio of any two is the number that measures the interval between the corresponding notes. These we shall call the "proportional numbers" of the scale. The next line contains the actual frequencies of these notes according to the most common method of tuning orchestral instruments ("International" or Low Pitch). These numbers actually vary slightly in different orchestras but their ratios remain the same. It may be noted that these numbers are equal to the proportional numbers multiplied by $10\frac{3}{4}$. The last line contains the frequencies of these notes as it has been for a long time the custom to use them in scientific work. These are equal to the proportional numbers multiplied by $10\frac{2}{3}$.

Some other intervals in the scale are named from the order of the notes reckoned from the first. Thus the interval from the first to the fifth, that is, from *C* to *G*, is called an interval of a *fifth*. From the proportional numbers we find that an interval of a fifth is equal to 36/24 or 3/2. It will readily be seen that the intervals from *E* to *B* and from *F* to *c* are also intervals of a fifth (3/2). The interval *C* to *F* is called an interval of a fourth and is equal to 4/3, as are also *D–G*, *E–A*, *G–c*. The intervals *C–E*, *F–A*, *G–B* are intervals of a *third* equal to 5/4. *C–D*, *D–E*, *G–A*, *A–B* are all called whole tone intervals, for while they differ slightly the difference is only that between 10/9 and 9/8 and can hardly be detected by most ears. The intervals *E–F* and *B–C* are much smaller and are called semi-tone intervals. For musical purposes intermediate notes named from these same letters but "sharped" or "flatted" (the black keys on a piano) are used, but further information on this subject belongs to the study of music.

598. Quality.—Two musical sounds of the same pitch and loudness may seem to the ear to be quite different. If the note *C* be sounded on a violin and on an organ, and be also sung, the ear will recognize the source at once. This difference we call a difference in quality.

On what does quality depend? We had no difficulty in connecting loudness with the amplitude of the wave-motion and

pitch with the wave length. Evidently quality must depend on the only other property in which air-waves differ, namely, the wave-form as shown by curves representing the waves. Fig. 465 shows three waves of the same amplitude and length but of different forms. The upper is what we have called a simple harmonic wave and is the form of wave given off by a tuning-fork or an organ-pipe when sounded very softly. The second is more like the wave from a violin string and the third like the wave from the human voice when singing the vowel *ah* (Miller). But voices differ greatly, and two voices singing the same note produce waves of somewhat different form. The same is true of waves emitted by a violin.

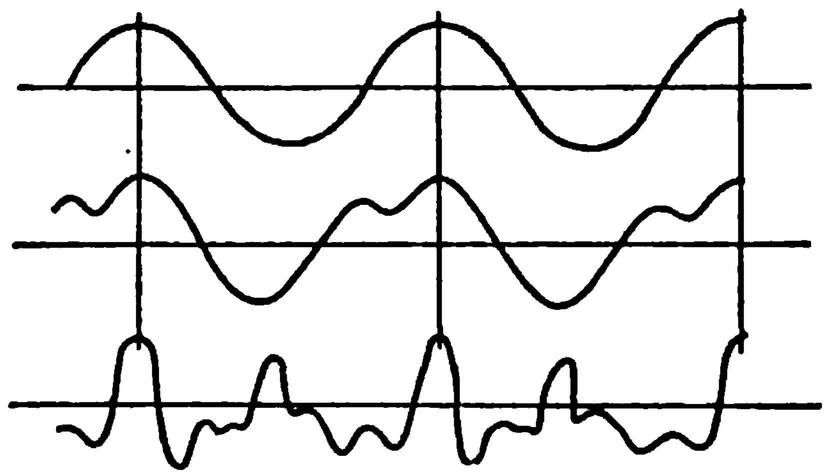


FIG. 465.—Different forms of waves.

599. Harmonics.—Notes with frequencies 2, 3, 4, etc. times the frequency of another note are called harmonics of the latter. Thus if the frequency of C is 256, the *first harmonic* of C is c which has a frequency of 512 and is an octave above C. The *second harmonic* of C is a note with a frequency of 768 and since 768/512 is the same as 3/2, the second harmonic of C is a fifth above c, that is, it is g. The third harmonic of C has a frequency of 1024 and since 1024/512 is the same as 2/1, it is an octave above c or two octaves above C. Hence it is c'.

Briefly stated, notes whose frequencies are $2N$, $3N$, $4N$, etc., are called the harmonics of the note whose frequency is N and the latter is called the *fundamental* of the harmonics.

If we now sound together such a series of notes, either with tuning-forks or on a piano, it will be found that (up to the seventh in the series) they combine together so harmoniously that it may be difficult to hear them separately. Thus the term harmonic is very appropriate.

600. Overtones.—If the sound from a large bell be listened to attentively, it will be found that it consists of several sounds of different pitch. The deepest of these is called the fundamental of the bell and the others are called *overtones* of the bell. The intervals between these overtones may be quite different on

different bells. In some cases the combination is such that the sound of the bell is pleasing. Other bells give an unpleasing or harsh combination. A note produced on a violin by a beginner is apt to consist of an unpleasant combination of a fundamental and overtones, while an accomplished violinist has learned how to produce an agreeable combination.

The difference between the terms harmonic and overtone should be noted. Harmonics refer to notes however they may be produced. When we are given the frequency of a fundamental we can at once calculate the frequencies of its harmonics by multiplying by 2, 3, 4, etc. Overtones refer to a particular instrument. An overtone of an instrument may not be a harmonic of the fundamental, but we shall see that the overtones of some instruments are harmonics of their fundamental tones.

601. Elementary and Compound Sounds.—Most people can without any special training distinguish a single sound from a mixture of sounds. Thus a tuning fork emits a single sound when it is struck gently. When it is struck violently, or when two tuning forks are sounded at the same time, the ear can usually tell us that a mixture of sounds is heard.

A well-trained ear is capable of doing more than this. For example, it can detect that, when a single piano string is struck, the sound produced is not single but consists of a fundamental and various harmonics. Thus the ear can do for musical sounds what the chemist can do for chemical compounds, namely, resolve them into their elements. The simple or elementary sounds of music are those in which the ear can detect no mixture of sounds. These are called *pure tones.* This at once suggests another question. What is the difference of the wave form of a pure tone and that of a mixed note, consisting of a fundamental and harmonics?

Now we have already seen that there are means by which a vibrating body can be made to record the form of its vibrations (§595). By these it is found that *the vibrations of a body emitting a pure tone are simple harmonic and the sound waves it emits are simple harmonic waves.* We have also seen that simple harmonic waves of different length when added together produce complex waves which may differ markedly from the simple harmonic form. Thus harmonics present with a fundamental alter the form of the

wave, and it is this wave form that determines the quality of the sound heard.

602. Beats Between Sounds.—When a white key and an adjacent black key near the bottom of the keyboard of a piano are struck at the same time, a distinct throbbing of the sound can be heard. The throbs are slow and can almost be counted. When the same is tried higher up on the keyboard, the throbs are more rapid. Throbs produced by two sounds of nearly the same pitch are called *beats.*

If two tuning forks of nearly the same pitch and mounted on their sound boxes be thrown slightly out of unison by attaching a small piece of wax to a prong of one, and if they be then sounded strongly by a bow, beats slow enough to be counted will be heard. When a larger piece of wax is used the difference of pitch of the forks is increased and the frequency of the beats also increases. In fact in all cases, *the frequency of the beats between two notes is equal to the difference of the frequencies of the notes.*

Since beats are due to two wave trains, of different wave lengths, coming to the ear at the same time, they are the result of what we have called interference of waves (§247). Fig. 152 may be taken as representing the production of beats between two trains of sound waves.

603. Nature of Vocal Sounds.—A vowel is a sustained sound of a variable pitch produced by holding the vocal cords and the resonating cavities of the mouth and throat in definite configurations, while consonants are explosive noises produced by the changes in the vocal organs preceding or following a vowel sound. Vowels uttered by the same voice at the same pitch differ in quality. Fig. 465 shows a record obtained by Miller for the vowel *ah* in father at a pitch of 182 by means of his phonodeik (§587). The precise cause of the difference of quality of different vowels has been a matter of long dispute. The first and natural supposition was that a certain vowel consisted of a definite combination of a fundamental and its harmonics in fixed proportions, whatever the pitch of the fundamental might be. If so the curve of Fig. 465 would be of the same form (for the same voice) whatever the pitch of the fundamental might be. This is contrary to results obtained by Miller and others. Helmholtz maintained that in a particular vowel sound there was a definite overtone characteristic of the vowel and having the same frequency no matter what the pitch of the fundamental might be. Miller found that at whatever pitch (between 129 and 259) the vowel *ah* was sounded, 60 per cent. of the energy was concentrated in an overtone of about 920 which was nearly but not quite constant. When a phonograph

is driven at less than its normal speed, the quality of a vowel in a singer's record is altered. This is contrary to the first view, since the change of speed does not change the *relative* pitch and energy of the combination of fundamental and overtones.

604. Difference Tones.—When beats between two tones are sufficiently rapid they coalesce and form a distinct beat tone, which may, however, be very weak. The frequency of the beat tone is the difference of the frequencies of the separate tones and when it is distinctly audible it is usually much lower than either component. When careful attention is paid, the beat tones between two piano strings or two violin strings can be clearly heard and faint though the sounds may be they have a distinct effect on the musical quality for trained ears. Beat tones may, if of too low a pitch to awake resonance in the instrument, which is the case when the beating tones are on the lower strings of a violin, be an effect purely in the ear, while the instrument itself may resonate to higher beat tones.

SOURCES OF MUSICAL SOUNDS

605. Musical Instruments.—The number of musical instruments is so great that only a few can be referred to here. They may be most simply classified according to the kind of body that is started vibrating in order to produce the sound.

Vibrating cords are used in violins, pianos, harps, etc. In an orchestra these are called *stringed instruments.*

Vibrating columns of air are used in organs, flutes, clarinets, etc. These are called *wind instruments.*

Vibrating rods, plates, bells, and membranes are used in what are called *percussion instruments* because they are sounded by striking.

606. Vibrations of Cords.—We have already considered cases of vibrations on cords when the vibrations are slow enough to be followed by the eye (Figs. 137, 160). When such waves are continually reflected from both ends the cord can divide up into vibrating segments of stationary waves. Each such segment is half of a wave-length in length. We shall now consider vibrations on a cord when they are rapid enough to produce sound waves, but there is no essential difference between the two cases.

When a cord stretched between two supports vibrates, it moves to and fro between two opposite extreme positions. The forms of the cord in the two extreme positions can be found by photography and in other ways and they are found to depend on the way in which the cord is started. The simplest case is when

the cord is very gently bowed at the middle. In this case the form is that of half of a simple harmonic wave, and, as might be expected, the note produced is a pure or elementary tone. The wave-length is therefore $2l$, where l is the length of the cord. Now the velocity of waves equals the product of frequency and wave-length or $v=n\lambda$. Hence

$$n=\frac{v}{2l}$$

and since the value of v is (§250) $\sqrt{T/m}$, T being the tension and m the mass of unit length of the cord,

$$n=\frac{1}{2l}\sqrt{\frac{T}{m}}$$

From this we see that the frequency is inversely proportional to the length. By shortening the "string" a violinist produces a higher note.

The frequency is also proportional to the square root of the tension. To tune a violin string to the proper pitch the tension is increased or decreased by turning a peg on which one end of the string is wound.

Finally the frequency is proportional inversely to the square root of the mass of unit length. A violin has four strings, the thickest being used for the lowest notes and the thinnest for the highest notes.

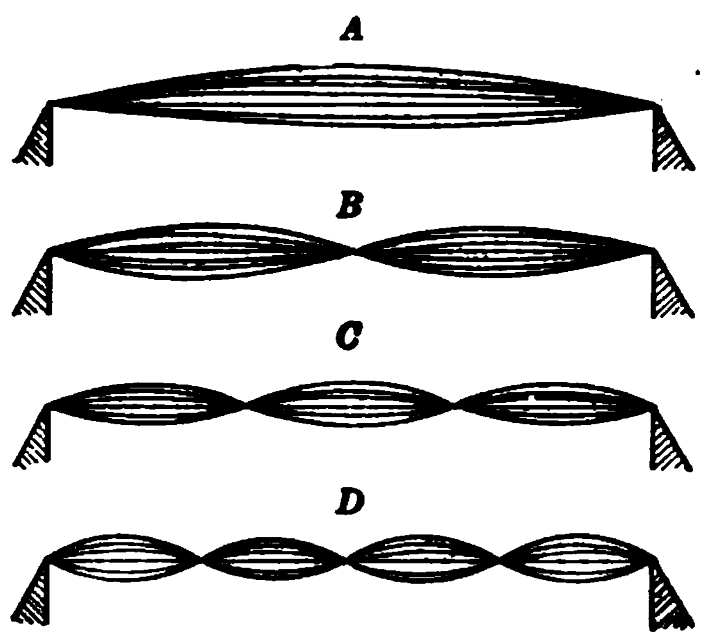


FIG. 466.

The vibration of a violin string is maintained by the bow, which alternately grips and slips on the string. The sound is not emitted directly by the strings bu by the resonating body of the instrument.

607. Overtones of Cords.—A cord can also divide into 2, 3, 4, etc., vibrating segments, as we have already seen in Fig. 137 and this is true whether the vibrations are too rapid to be followed by the eye or not. To produce vibrations with two segments on a cord touch it lightly at the center and bow one-half (Fig. 466, *B*). It will be seen that both halves vibrate and the ear

will hear a note higher by an octave than the fundamental. This is the first harmonic of the fundamental. If the finger be one-third of the length of the string from one end, and either the shorter or the longer part of the string be bowed, the note will be the second harmonic of the fundamental, and so on. In this way the violinist can produce a great variety of harmonics on the various strings.

Thus the overtones of a cord are harmonics of the fundamental. To explain why this should be so let us suppose that in any case the number of segments is N. Then if l is the length of the cord, the length of each segment is l/N and the wave-length is therefore $2l/N$. From this and the general relation $v=n\lambda$, we get

$$n=\frac{v}{\lambda}=\frac{vN}{2l}$$

Now l is the constant length of the cord and v is also constant so long as the tension of the cord is unchanged. Hence if we give values 1, 2, 3, to N we see that the frequencies are as 1, 2, 3, etc.

608. Complex Vibrations of Cords.—While we have described separately the different ways in which a cord can vibrate they can in reality take place at the same time. When a cord is bowed at one-fourth of its length from one end, a good musical ear can detect both the fundamental and the first harmonic. Or if, while it is vibrating in this condition, it be lightly touched at the middle, the fundamental will cease, but the first harmonic will be heard to continue for a moment. In a similar way we can show that other harmonics are present.

The particular overtones that are present in the complex vibrations of a cord depend chiefly on where it has been struck or bowed, for it is evident that no form of vibration that would require the point struck to be a node can be present. Some of the possible combinations are more pleasing than others. The most pleasing combination seems to be when the cord is struck at about one-eighth from one end and this is the actual practice in the piano and, to some extent, in the bowing of the violin also.

The explanation of the coëxistence of different forms of vibration depends on the fact that waves of different length can travel along the cord at the same time. Each such train, being reflected

at the ends, produces its own system of nodes and vibrating segments.

The form of the resultant wave on the string evidently depends on the particular combination of vibrations present, and for each combination there is a distinct characteristic quality of the complex sound heard. This is another proof that the quality of a sound depends on the form of the wave that causes it.

609. Vibrations of Air Columns.—Air in a tube open at one or both ends can be made to vibrate and emit a musical sound by blowing across an open end. Blowing with different degrees of strength will produce notes of different pitch. Gentle blowing will produce the lowest note and we shall consider this first.

Let us first suppose that the tube is closed at one end and for brevity let us call this a stopped tube. The vibrations of the air column are stationary waves (Fig. 467,*a*). Compressions started by the blowing at the open end travel to the closed end and are there reflected, so that the motion at any point in the tube is due to the superposition of two trains of waves travelling in opposite directions. The closed end is a place of no motion and is therefore a node of the stationary waves. The open end is a place of the greatest freedom of motion and is therefore the middle of a loop or vibrating segment. Now the distance from a node to the middle of a vibrating segment is one-fourth of a wave-length, λ. Hence, if l be the length of the tube,

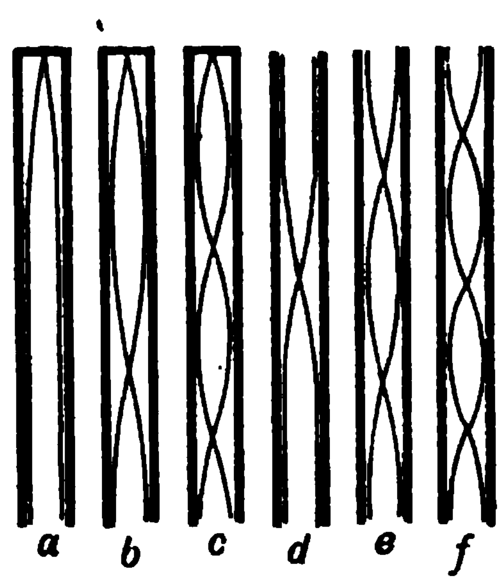


FIG. 467.

$$\tfrac{1}{4}\lambda = l \qquad \text{or} \qquad \lambda = 4l$$

If the tube is open at both ends, it is found that there is a node at the middle (Fig. 467*d*), while the ends are middles of loops. Hence it is readily seen that

$$\tfrac{1}{4}\lambda = \tfrac{1}{2}l \qquad \text{or} \qquad \lambda = 2l$$

From this we see that a stopped tube and an open tube of the same length produce waves with lengths as 2 to 1 and the notes have therefore frequencies as 1 to 2, since frequency varies inversely as wave-length. The note of the stopped tube is therefore an octave below that of the open tube.

Strictly speaking an open end is not exactly the middle of a loop, for a condensation returning to the open end does not reach full freedom of expansion until it has passed out a short distance. It has been found that the middle of the loop is in reality about .6 of the radius of the tube beyond the open end.

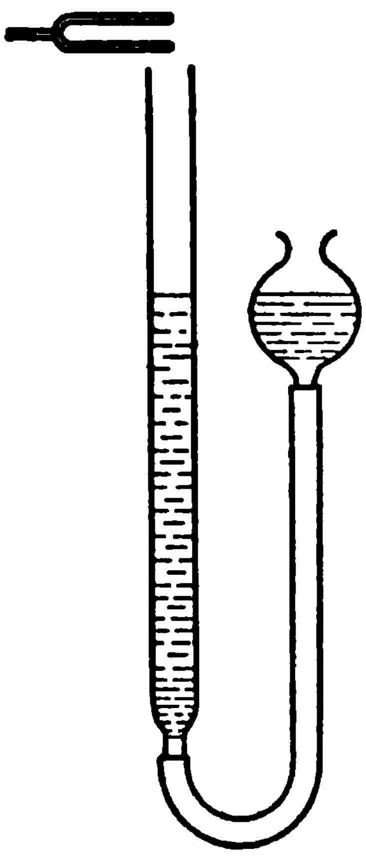


Fig. 468.—Resonating tube.

610. Resonance of Air Columns.—If over the open end of a stopped tube we bring a vibrating tuning fork of the same frequency as the lowest note which the tube will emit, resonance, or response of the tube to the fork, will take place and the note of the fork will come out strongly. A tube of variable length can be tuned exactly to the pitch of the fork and for this purpose a tube closed at the lower end by a column of water that can be raised or lowered is suitable. An open tube with an extension piece that slips over the end of the tube can be tuned to a fork. If a stopped tube and an open tube be tuned to the same fork, the former will be half as long as the latter.

The interaction between the fork and the tube is similar to that between the hand and a swing when the latter is being started. The forward and backward motions of the hand must be timed to agree with the corresponding natural motions of the swing. Similarly the fork must complete one vibration in the

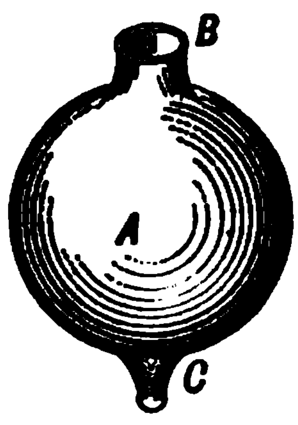


Fig. 469.—Spherical resonator.

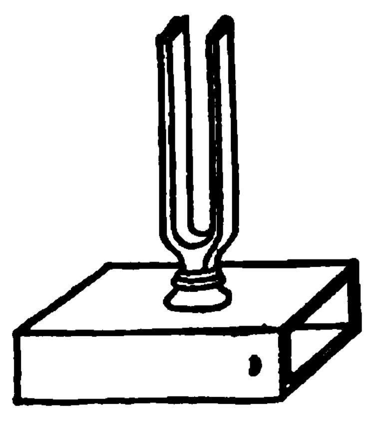


Fig. 470.—Tuning-fork on resonating box

time of one vibration of the air column, that is, in the time in which the sound travels one wave length, which, as we have seen is four times the length of the stopped tube and twice that of the open tube.

Other forms of air cavity can also resonate and emit definite

notes. Spherical resonators have been much used in analyzing complex sounds. For many purposes it is convenient to mount heavy tuning forks on resonating boxes, each of the same fundamental pitch as the fork. While the fork and the box agree in their fundamental tones, they differ as regards their overtones and the note emitted is much louder and purer than the fork alone can emit.

611. Organ Pipes.—An organ pipe is a tube the air in which is maintained in vibration by a jet of air from a wind chest. The jet is forced through a narrow slit at the mouth of the pipe and strikes against a sharp edge which borders an opening on one side of the pipe. The farther end of the pipe may be open or stopped, but the end at the mouth must be regarded as open. The first effect of the jet is to start either a condensation or a rarefaction in the pipe, depending on whether the jet is on the whole directed more to the inner or the outer side of the sharp edge. Thereafter the course of events is determined by the return of the pulse after reflection from the farther end. The arrival of a condensation forces the jet outward and it then by its suction reinforces the rarefaction which succeeds the condensation and travels up the pipe. When a rarefaction arrives the jet is forced inward and reinforces the condensation which follows. Thus the pitch is controlled by the natural period of the pipe and the energy required to sustain the vibrations is derived from that of the jet.

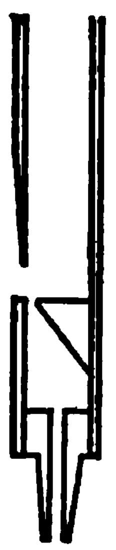


FIG. 471.

A stopped pipe is tuned by adjusting a movable plug which closes the stopped end. An open pipe usually has a small hole near the open end, and the pipe is tuned by adjusting a small strip of metal that partly closes the hole.

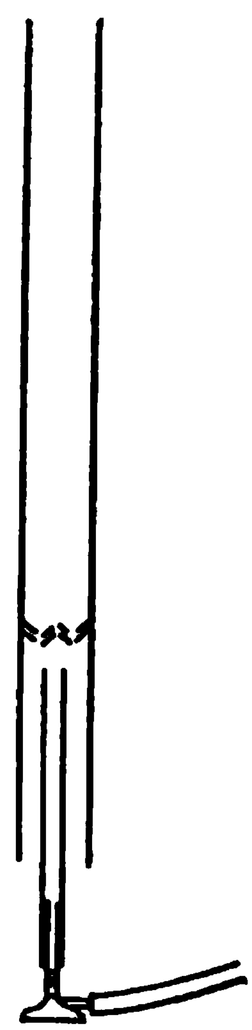


FIG. 472.

An interesting illustration of another method of supplying the energy of vibration of a sounding tube is shown in Fig. 472. A long thick metal tube has a sheet of wire gauze inserted to about one-fifth of its length from one end. If the gauze be heated by a Bunsen burner (with a long extension tube), the tube will sound very loudly when the burner is removed. The whole explanation

is complex, but the supply of energy is readily explained. A condensation coming to the heated gauze removes heat more rapidly than a rarefaction does. Hence the energy of each condensation is reinforced by the supply of energy received from the heated gauze.

612. Tones of a Stopped Organ Pipe.—The closed end of a stopped pipe is always a node and for the fundamental tone it is the only node, the open end being the middle of a loop. The first overtone has a second node, the open end being still the middle of a loop (Fig. 467, *b*). Hence the second node is one-third of the length of the pipe from the open end. Thus the wave length, which is always four times the distance from the middle of a loop to the nearest node, is 4/3 of the length of the pipe. For the second overtone there are three nodes and so on. Thus it is readily seen that if l be the length of the pipe and N the frequency of the fundamental the series of wave-lengths and notes are

$$\begin{array}{llll} 4l, & 4l/3, & 4l/5, & \ldots \\ N, & 3N, & 5N, & \ldots \end{array}$$

It will be noticed that the overtones are harmonics of the fundamental, but those of frequency $2N$, $4N$, etc., are not produced by a stopped pipe.

Except when blown very softly a pipe produces one or more overtones along with the fundamental and the number and intensity of these determine the quality of the complex sound. The absence of the harmonic $2N$ gives the stopped pipe its characteristic somewhat dull tone.

613. Tones of an Open Organ Pipe.—Both ends of an open pipe are mid-points of loops. At least one node is required for stationary waves, and, when an open pipe sounds its fundamental tone, this node must evidently be at the middle of the pipe. For the first overtone there are two nodes, one at one-fourth of the length of the pipe from one end and the other at an equal distance from the other end (Fig. 467, *e*). For the second overtone there are three nodes and so on. Thus it is readily seen that the wave-lengths and frequencies are

$$\begin{array}{llll} 2l, & 2l/2, & 2l/3, & \ldots \\ N, & 2N, & 3N, & \ldots \end{array}$$

The brighter quality of the open pipe is due to the presence of the overtone $2N$.

614. Longitudinal Vibrations of Rods.—A rod clamped at its middle point can be made to vibrate by stroking it with a rosined glove. Since both ends are free and the middle is fixed, the fundamental vibrations are similar to those of an open organ pipe, and the wave-length is twice the length of the rod. The overtones follow the same law as those of an open organ pipe. The first overtone may be produced by clamping the rod at points one-fourth of the length from the ends (as in Fig. 476) and so on.

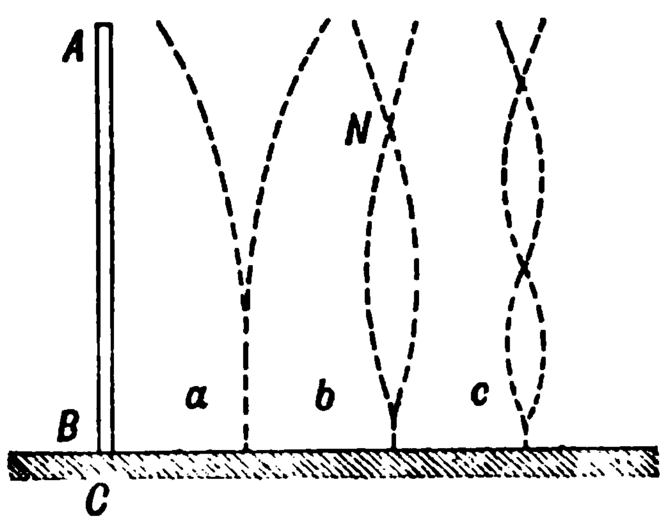


FIG. 473.

615. Transverse Vibrations of Rods.—A rod or strip of metal clamped at one end can be readily set into transverse vibrations by a blow or push at the free end. Of two rods of the same kind the shorter vibrates more rapidly than the longer, and if the vibrations are rapid enough they produce a musical note. The fixed end is, of course, a node, but overtones with

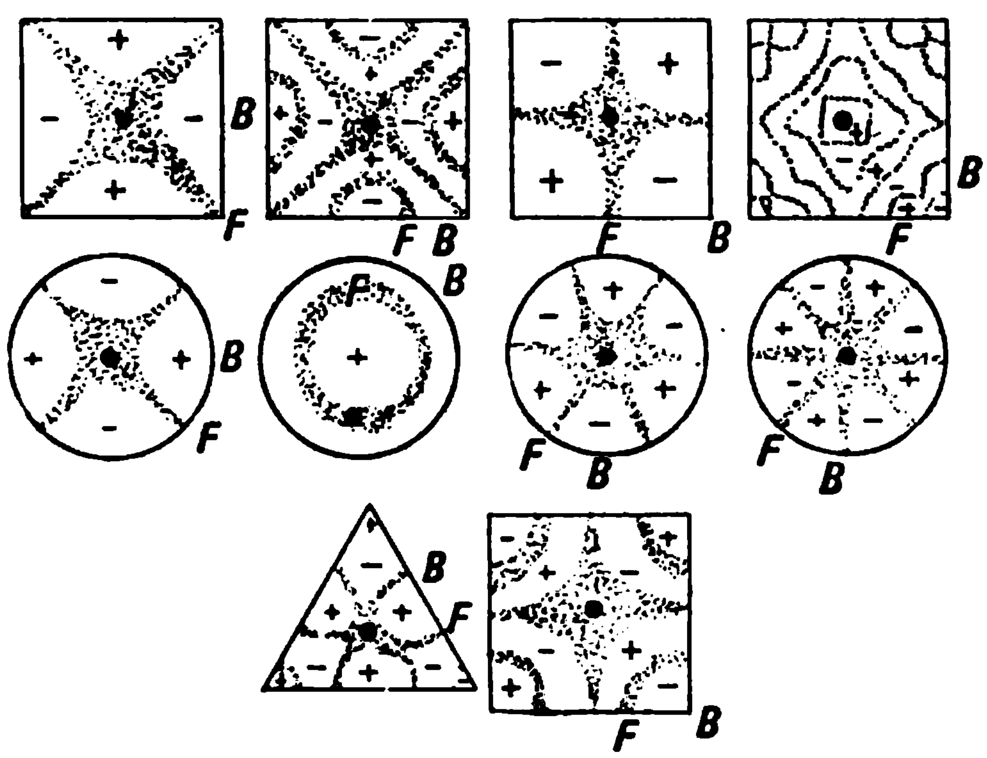


FIG. 474.

two or more nodes can also be produced. The frequencies of these are not in simple ratios to the frequency of the fundamental. Hence the overtones are not harmonics. Such vibrations of strips of wood or metal are used as *reeds* in many musical instruments, such as the clarinet, oboe, bassoon, and reed organ. The reed is placed at the end of a pipe that resonates to the vibrations of the reed.

A rod clamped at the middle will vibrate transversely with a node at that point. A tuning fork consists essentially of such a rod with the ends turned parallel. The overtones of a tuning fork are not harmonics of the fundamental, but if the fork is sufficiently thick and is not struck too violently, the overtones, besides being very high are so weak as not to produce disagreeable effects.

Large metal tubes vibrating transversely are sometimes used in operatic music to imitate deep-toned bells.

616. Vibrations of Plates.—A square or round plate of metal, the middle of which is screwed to the end of a rod, can be made to vibrate in a great variety of forms by stroking with a bow. The nodes or lines of no motion are beautifully shown by fine sand on the plate. In the fundamental mode of vibration there are only two of these lines, but in the higher modes of vibration a large number, straight or curved, appear.

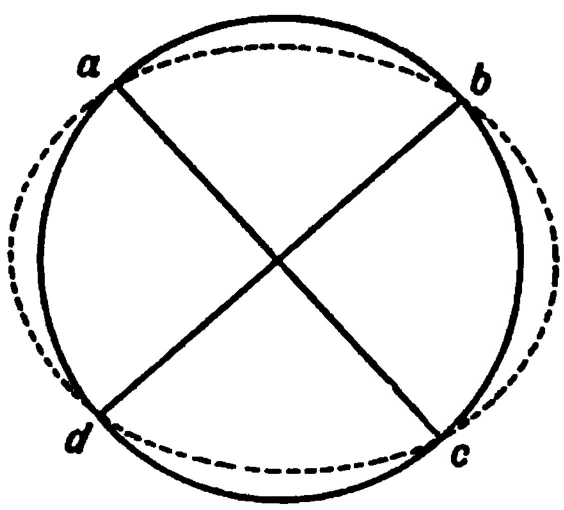


FIG. 475.

617. Bells.—A bell may be thought of as a round plate turned up to a cup shape. The fundamental mode of vibration is like that of the plate, with two nodal lines; but when the bell is struck, many overtones are pruduced along with the fundamental, and some of these are so much stronger than the fundamental that the latter frequently cannot be heard. In fact it is usually the fourth overtone that is loudest and is taken by the ear as giving the pitch of the bell.

VELOCITY OF SOUND. EXPERIMENTAL METHODS

618. Resonating Tube Method.—From the length of a tube that resonates to a tuning fork of known frequency the velocity of sound in the gas in the tube can be found by means of the relation $V = n\lambda$. For this purpose a glass tube open at one end and containing an adjustable column of water is convenient. From the account of the vibrations of a stopped organ pipe (§ 612) it is seen that, if the tube be of sufficient length, it can resonate when the length of the column of gas is $\lambda/4$, or $3\lambda/4$ or $5\lambda/4$, etc., where λ is the wave length of the sound of the tuning fork. When it resonates with more than one node the distance between two

consecutive nodes is accurately $\lambda/2$, while the distance from the upper node to the open end is only approximately $\lambda/4$. Hence the best value of the wave-length and velocity is obtained from the former.

619. Kundt's Dust Tube Method.—This method differs from the preceding in that a metal rod vibrating longitudinally is used instead of a tuning fork and, since the pitch is very high, the air column divides into numerous vibrating segments. These segments are clearly shown by cork dust which gathers at the nodes. One end of the rod projects into the tube and carries a light disk of slightly smaller diameter than that of the tube. It is convenient for stability to clamp the rod at two points, each at one-fourth of the length of the rod from one end. The

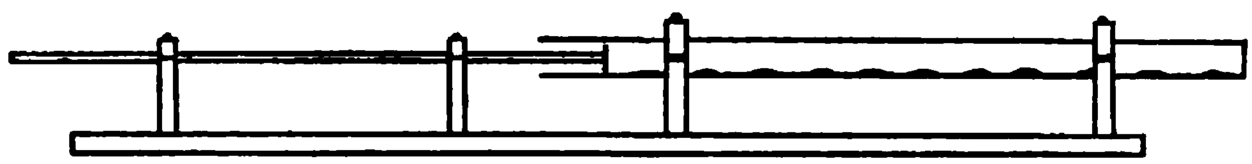


FIG. 476.

wave-length of the waves in the rod is then equal to the length, L, of the rod. The wave-length of the sound in the gas is twice the distance between two adjacent dust heaps or $2l$. Let V be the velocity of the waves in the rod and v the velocity of sound in the gas. Then if n be the frequency of the vibration, which is the same for both, $V=nL$ and $v=2nl$.

Hence

$$\frac{v}{V}=\frac{2l}{L}$$

If we assume the velocity of sound in air to be known we can, from this relation, find the velocity of the waves in the rod and also the velocity of sound in any other gas introduced into the tube. The method also lends itself to the study of the effect of change of temperature on the velocity of sound and it has been modified so as to make it suitable for finding the velocity of sound in a liquid.

ACOUSTICS OF HALLS

620. Defects in the Acoustics of Halls.—In some halls the "hearing" is satisfactory, in others unsatisfactory. A full treat-

FIG. 477A.—(Sabine, *American Architect*.)

FIG. 477B.—(Sabine, *American Architect*.)

ment of the subject is impossible here, but a few points may be considered briefly.

Sound reaches an auditor in a hall not only by waves coming directly from the source, but also by waves reflected by walls (including in this term all reflecting surfaces) and the paths usually differ in length. The result may be an increase in loudness but more or less confusion of sound also follows. A speaker utters two or three syllables per second and a musical instrument may emit several notes in a second and reflection causes an overlapping of the separate sounds. The two chief defects resulting may be classified as *echo* and *reverberation*. A separate echo is due to some particular reflecting surface and may be reduced by alterations in the surface. Reverberation consists in a prolongation of the sound in all parts of the hall due to repeated reflections from all walls, somewhat as a room (with white walls) may be illuminated throughout by light entering by a single small window. Reverberation is always a drawback for speech but it cannot be eliminated without reducing loudness to an undesirable extent. For music a certain amount of reverberation is an advantage, but a larger amount is detrimental.

Fig. 477 shows the history of a single wave of sound started from the stage of a theater and reflected by various surfaces. In *A* we have the direct wave .07 of a second after it has started, together with several reflected waves that will cause weak echoes. *B* shows a part of the direct wave .07 of a second later and a complex system of reflected waves. The figures are from photographs obtained by Professor W. C. Sabine, to whom we owe most of our knowledge of Architectural Acoustics, by the method referred to in §588, a small model of the theater being used for the purpose.

621. Remedy for Reverberation.—If the walls of a hall did not reflect sound there would be no reverberation. When it is present to an undesirable extent its amount may be decreased by covering some part of the walls with soft materials that reflect little sound. It is usually not a matter of importance on what wall or part of a wall the deadening material is placed. This is due to the fact that the sound starts out from the source in all directions and travels about 1140 ft. in a second. This distance is so great compared with the dimensions of an ordinary hall that the rays are in general reflected many times in a second and as

much approximately falls on each square unit of area wherever it may be situated. Yet there are places in some halls where more or less than the average amount of sound reaches the walls and for such special cases a special study of the hall is necessary. For the details of this we must refer to Sabine's papers.

The reader may have heard of cases in which an attempt has been made to reduce reverberation by stretching fine wires across a hall. These and other similar devices are entirely useless, as has been shown by careful study of the results.

622. Duration of Residual Sound.—Since reverberation is due to the continuance of a sound after it has been produced, the extent of the reverberation in a hall can be estimated by measuring the length of time a sound is heard in the hall after the source has ceased emitting waves. Sabine studied this question by means of an organ pipe having a frequency of 512 and blown under a definite pressure, together with a chronograph for recording the length of time the residual sound was heard. He found that the residual sound was heard for a measurable number of seconds and died away according to an exponential law. From the data obtained it was possible to calculate the time required for any specified decrease in intensity, and it was convenient to define "the duration of residual sound" in a hall as the time required for the intensity to fall to one millionth (10^{-6}) of its original value. This we shall denote by T. While the value of T is different for notes of different pitch and such variations were studied by Sabine, it will be convenient in this brief account to restrict it to the particular frequency (512) referred to, for it was found that, for orchestral music, T, thus defined, should be about 2.3 seconds. A hall for which T is as great as 3 seconds or as small as 2 seconds is not satisfactory for orchestral music.

623. Coefficients of Absorption.—An open window may be regarded as a non-reflector, that is a perfect absorber, since all the sound which falls on it passes out. Sabine found by experiments on the duration of residual sound in an auditorium what area of any particular drapery would cause the same absorption of sound as 1 sq. m. of open window, and therefrom he deduced the absorbing power of the material. If, for example, it took 3 sq. m. of a material to produce the same decrease of the residual sound as 1 sq. m. of open window, each square

meter of the material must have absorbed one-third as much as the square meter of open window. But the latter absorbed all of the sound that fell on it. Hence the drapery absorbed one-third of the incident sound, therefore its coefficient of absorption was ⅓. Sabine found by such experiments the absorbing power of an average auditor and of various articles which are commonly found in a hall. Some of these coefficients are given in the following table. For additional data Sabine's papers must be consulted.

COEFFICIENTS OF ABSORPTION

Open window	1.000
Wood sheathing (hard pine)	.061
Plaster on lath	.034
Plaster on tile	.025
Brick set in Portland cement	.025
Carpet rugs	.20
Shelia curtains	.23
Hair felt (2.5 cm. thick, 8 cm. from wall)	.78
Audience (per person)	.44

624. Total Absorbing Power of a Hall.—Each surface absorbs in proportion to its area and coefficient of absorption. If, therefore, we multiply each area of surface of a particular kind by its coefficient of absorption and add the products we shall get the total absorbing power, a, of the auditorium, or

$$a = a_1s_1 + a_2s_2 + \ . \ . \ . \ = \Sigma as$$

To take account of the audience in this sum we multiply the number supposed to be present by the absorbing power per person.

625. Effect of Size and Shape of Hall.—Sabine found that in general the shape of a hall of given total volume and given total absorbing power had, to a first approximation at least, no effect on the length of duration of residual sound in the hall. In fact, other things being equal, if the sound diffuses as rapidly as we have supposed to all parts of a hall and so becomes nearly uniformly distributed throughout, the absorption must depend only on the average frequency of reflection of the rays starting from the source, and it can be shown theoretically that this average frequency depends only on the volume.

626. Sabine's Formula.—Sabine found by experiments on many halls that the length of duration of residual sound could be expressed by

$$T = .164 \frac{V}{a}$$

where V is the total volume of the hall in cubic meters and a is its total absorbing power, areas being calculated in square meters. Strictly speaking the velocity of sound should appear in the denominator but it is assumed to have the average value of 342 m/s and is included in the factor .164.

The above formula can be applied at once to calculating the duration of residual sound in an auditorium, existing or contemplated, and thus its suitability for music, and approximately also for speech, can be tested. It can be applied also to remedy the defects in an existing hall by suitably varying the value of a. From the formula we can also derive immediate answers to such questions as the effect of the presence of the audience on the "hearing" in a hall, the result of copying an existing satisfactory hall but on a larger scale and so on. These may be left as exercise to the reader.

References

RAYLEIGH'S *Theory of Sound* is a standard work containing both theory and experiment.

HELMHOLTZ'S *Sensations of Tone* is a classical work of great originality.

POYNTING AND THOMSON'S *Sound.*

POYNTING on *Sound* in the Encyclopedia Britannica.

BARTON'S *Sound.*

These three works contain clear statements of experimental facts with considerable theory.

MILLER'S *Science of Musical Sounds* is an interesting and original work, copiously illustrated, on the experimental analysis of sound waves.

SABINE'S articles on *Architectural Acoustics* (American Architect, 1900, 1913; Proc. American Academy of Arts and Sciences, 1906) are of great originality and of fundamental importance.

Problems

1. What is the ratio of the velocity of sound on a hot summer day (35°C.) to that on a cold winter day (– 20°C.)?
2. A Band is playing on a steamer which is travelling with a speed of 20

miles per hour. With what speed must an observer move in the opposite direction in order that every note may be lowered by a semitone?

3. Find the pitch of the fundamental and of the first two overtones of a stopped pipe 1 m. long at 16°C.
4. What is the length of an open pipe, if the pitch of its first overtone is 512?
5. A tuning fork makes 2 beats per second with a standard fork the frequency of which is 512. When a small piece of wax is placed on a prong of the former the number of beats is decreased. What is its pitch?
6. A tuning fork of frequency 384 makes 2 beats per second with a vibrating string. In what proportion must the tension of the string be changed in order that the two may be in unison?
7. A stopped pipe resonates to a tuning fork, the pitch of which is 258, when adjusted to a length of 32 cm. and also when the length is 98 cm., the temperature being 18°C. What is the ratio of the specific heats of air?
8. A glass rod is used in Kundt's experiment. If the dust piles are 5 cm. apart and the distance between the clamped points on the rod is 70 cm., what is Young's modulus for this glass, the density of glass being 2.7 gm./cu. cm.?

LIGHT.

By E. Percival Lewis, Ph. D.

Professor of Physics in the University of California.

GENERAL PROPERTIES.

627. Radiation.—As in the cases of Heat and Sound, the word Light has acquired two distinct meanings. The primary and more familiar one is that which is associated with the sensation of vision. Nearly all that relates to this aspect of the subject lies within the province of the psychologist. The physicist, however, generally uses the term in an objective sense, with reference to the external agencies which may excite the sensation of light if allowed to act on the eye. The visible radiation which affects a normal eye will also affect a photographic plate, a thermometer, or other sensitive detector of heat. It will be found, after analyzing the radiation from the sun, electric light, or other sources with a prism, that beyond the violet and the red lie non-luminous radiations which will affect a photographic plate or a thermometer, and it has been shown that the oscillations of an electric spark between metallic terminals are accompanied by the radiation of electric waves through space. There is, as we shall see, no fundamental qualitative difference between these various radiations, and it is due merely to a special property of the eye that some of them excite the sensation of light while others do not. This is analogous to the selective resonance of a piano wire, which will respond to certain notes and not to others. Just as some ears can detect sounds of such high pitch as to be inaudible to others, some eyes can detect ether radiations lying somewhat beyond the limits of perception of the ordinary eye. In the following pages the whole range of these radiations so far as they are known will be considered. As a matter of convenience, the term **Light**, which strictly speaking would apply only

to the radiations exciting the sensation of light, will be used in a figurative sense to include the entire range of radiations which are alike in their general properties, and which were once very artificially classified as luminous, actinic, and heat radiations.

628. Sources of Light.—The best known are the sun, the physical nature and condition of which are as yet not fully understood, solid bodies at a high temperature, such as the calcium light, electric arc and incandescent lights, and luminous flames. If a piece of cold porcelain is held over the flame of a candle, lamp, or gas jet, it will become covered with finely divided carbon, while no such deposit is observed in the case of a non-luminous Bunsen or alcohol flame. This suggests that the luminosity of these flames is due to the presence of incandescent carbon particles. This idea is strengthened by the fact that when the base of a Bunsen burner is closed the flame becomes luminous and smoky; when open, enough oxygen is admitted to combine with all the carbon set free by the dissociation of the coal gas, and the flame is then non-luminous. The carbon oxides formed are permanent gases, and there is no evidence that such gases can be made luminous by high temperature alone. Any gas may be made luminous, however, by the passage of an electric discharge through it, but this luminosity does not seem to be accompanied by very high temperature. There are, in fact, many cases in which light is emitted at a very low average temperature of the source. As examples may be mentioned the various types of phosphorescence, some of which are most active at temperatures as low as that of liquid air, the aurora due to electrical discharges through the highly rarefied and very cold upper atmosphere, and the light emitted by fire-flies and glow-worms.

629. Rectilinear Propagation.—One of the earliest observations concerning light was that it travels in straight lines, in a homogeneous medium. These lines of propagation or "rays" may be made to alter their direction only by one of two methods —by reflection, when they fall on the boundary between two media, or by refraction, when they pass obliquely from one medium to another, or through a medium of varying density.

630. Shadows and Eclipses.—Rays pass in straight lines by the edges of an obstacle, so that the space behind it is screened

from the light. If the latter comes from a very small or "point" source the shadow would be sharply defined if the propagation were strictly rectilinear; as a matter of fact, close observation shows in all cases that the light fades gradually into the shadow. This very significant fact proves that light travels only approximately in straight lines; there is always more or less lateral spreading. Strictly speaking, there is, then, no such thing as a ray of light, if we mean by this term propagation along a geometrical line. The explanation of this spreading will be given later (§701 *et seq.*).

A more obvious cause of the lack of sharpness in shadows is to be found in the fact that most sources of light are not even approximately points, but are of finite area. This gives rise to the distribution of light and shadow shown in Fig. 478 and Fig. 479. The first represents the shadow cast by an object larger than the source; the second, that due to an object smaller than the source; for example, the shadow of the earth due to the sun. In each case there is a region of complete shadow behind the obstacle, called the umbra, into which no light from any part of the source can enter. Around this there is a region called the penumbra, which receives light from a part of the source, the effective portion of the latter increasing in going outward from the umbra. When the moon lies entirely within the shadow cone of the earth it is said to be completely eclipsed; when it passes through the penumbra or partly through the umbra and partly through the penumbra it is partially eclipsed.

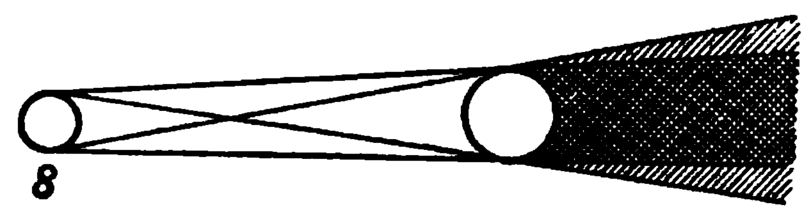


FIG. 478.

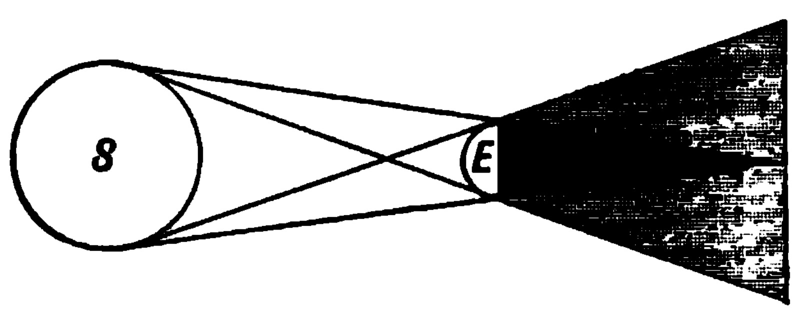


FIG. 479.

631. Parallax.—This well-known phenomenon depends upon the rectilinear propagation of light. By parallax is meant the apparent displacement of an object due to the real displacement of the observer. For example, if the observer moves from O_1 to O_2 (Fig. 480) A will appear to be displaced an angular distance $\alpha+\beta$ to the left with reference to B. That object which seems to be displaced in a direction opposite to the motion of the observer is evidently the nearer. To one traveling on a railroad train objects near at hand appear to be moving backward, those

at a distance in the same direction as the observer. If two objects are coincident in position or equally distant their relative parallax vanishes. This gives a useful method of finding the apparent position of the image formed by a lens or mirror, or of focusing the cross thread of a telescope. When the latter and the image of a distant object are both distinctly seen and have no relative parallax they are coincident in position and both in focus.

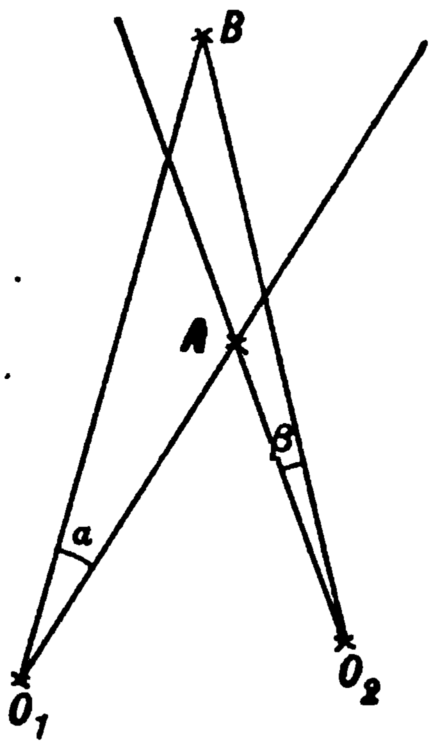


FIG. 480.

In astronomy horizontal parallax is defined as the angle subtended by the semi-diameter of the earth from any body of the solar system. Annual parallax is the angle subtended by the semi-diameter of the earth's orbit from the more distant fixed stars. The distance between the sun and the earth may be determined by observing the transit of an inferior planet, Venus for example, across the sun's disk. Observers at A and B (Fig. 481) note the instants at which Venus appears to enter the sun's disk as viewed from their respective stations. From the interval between these two contacts and the known angular velocity of Venus around the sun the angle α may be determined, and from that and the base line AB the horizontal parallax and distance of the sun may be calculated. Of course correction must be made for the motion of the earth between the instants of contact.

632. Pinhole Image.—Another effect of the approximately rectilinear propagation of light is the formation of an inverted image of the source by light passing through a small orifice such as a pinhole. If any source, for example, a candle, is placed opposite such a hole in a screen S_1 (Fig. 483) light from the point P will pass through the opening in a narrow cone or *pencil* and illuminate a small patch at P_2 on a screen S_2. Light from Q will form a small patch at Q_2,

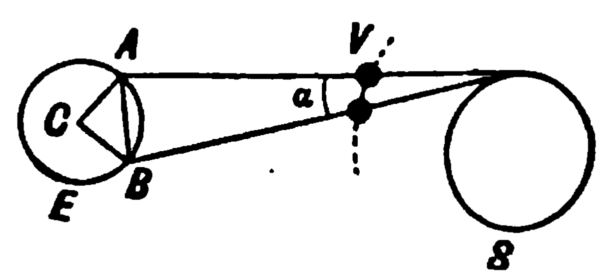


FIG. 481.

and light from any other point of the flame will fall on a corresponding point of the screen S_2. The group of patches will in form, color, and relative brightness reproduce the candle flame, but evidently inverted in position. The pinhole forms an image like that due to a condensing lens, but the total light in the

pinhole image will be less than that formed by the lens in the proportion of the area of the pinhole to that of the lens. As the image is due to a group of overlapping patches, it will not be so sharp in outline as that made by the lens. The blurring will increase with the size of the opening or when the source is brought near the screen, thus increasing the angle of the transmitted cone. The object and its image subtend equal angles at the pinhole, so that their linear magnitudes are in the same ratio as their respective distances u and v from the screen S_1. This is also true of images formed by any other optical device, such as a mirror or lens. Landscape photographs of great softness and beauty may be made by the use of the pinhole camera.

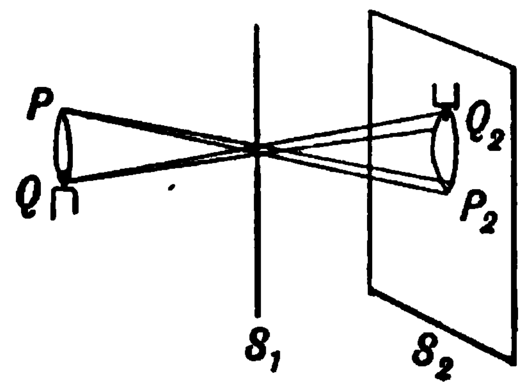


FIG. 482.

633. Reflection, Regular and Diffuse.—When light falls on a smooth polished surface it is reflected in a definite direction. This is called regular reflection. The plane including the direction of the incident light and the normal to the surface at the point of incidence is called the *plane of incidence.* The angle between the incident pencil and the normal to the surface is called the *angle of incidence;* that between the reflected pencil and the normal is called the *angle of reflection.* Experiment shows that (1) *the angle of reflection is equal to the angle of incidence;* (2) *the reflected pencil lies in the plane of incidence.* It is evident from the first law that if a mirror is rotated through a given angle about an axis perpendicular to the plane of incidence, the reflected pencil will be rotated through twice that angle.

When light falls on a rough unpolished surface it is reflected in all directions. This is called diffuse or irregular reflection. There is no essential difference between regular and diffuse reflection except that in the latter we may imagine reflection to take place from an infinite number of infinitesimal plane surfaces orientated in all directions.

634. Visibility of Objects.—On a clear night, where there is no moonlight, the stars and planets appear against a background of black sky. The space around the earth's shadow cone is filled with sunlight, but we do not see it unless it is reflected from some planet or the moon. If a beam of light is passed through a

vessel of distilled water its path is invisible. If a beam of sunlight enters a dark room it cannot be seen unless dust particles are floating in the air. A drop of milk in the water or a little dust stirred up in the room will cause the path of the light to flash out brilliantly. Such experiments show, as might be expected, that light does not excite the sensation of luminosity unless it enters the eye directly from the source or by reflection. Ordinary objects are visible because they reflect light diffusely into the eye, and they may be regarded as secondary sources of radiation. A perfect reflector would itself be invisible, all the light reflected from it appearing to come from the image of the source, not from the reflector.

635. Transmission and Absorption.—Light travels through some media, for example most gases, glass and water, with scarcely any appreciable diminution of intensity. Other media may transmit little or none, or certain colors only; such media are said to show *general* or *selective* absorption. In cases where absorption occurs there appears to be a loss of radiant energy, but it may be shown that it changed to other forms, usually heat (§330 *et seq.*).

636. Transparency, Translucency, Opacity.—Any substance which transmits a large fraction of the incident light without scattering it is said to be transparent. As indicated by this term, objects may be seen clearly through such substances. Objects which absorb all the unreflected incident light are said to be opaque, and act as perfect screens. Evidently any perfect reflector must also be perfectly opaque, but in this case opacity is not due to absorption. Substances differ widely in these properties, varying from almost perfect transparency to almost perfect opacity. The most transparent media known show some absorption, which increases with the length of path; hence any substance will become opaque if a sufficient thickness is taken.

No light penetrates to great depths in the ocean, although a layer of water of considerable thickness is transparent. On the other hand, light will penetrate to a slight depth in any medium, so that thin layers of metal or of carbon are found to be transparent. Some substances are selectively transparent; red glass will freely transmit red light, but not the other colors, and a thin sheet of hard rubber, which appears to be opaque, will transmit radiations lying a little outside the red of the spectrum.

Some substances transmit light, but scatter it so that objects cannot be

clearly seen through them. These substances are called translucent. The effect is caused by diffuse reflection within the medium, due to discontinuity or non-homogeneity of structure, as in the case of powdered glass, paper, or water containing finely divided particles. Some substances, such as paraffin, are homogeneous and transparent when in the fluid state, and translucent when in the solid state. The atter effect is apparently due to granulation or crystallization.

637. Refraction.—When light passes obliquely from one transparent medium to another a part is usually reflected, while that which enters the second medium changes its direction abruptly at the boundary. Generally (but not always) in passing from a lighter medium to a denser the light is deflected toward the normal to the boundary. This is called refraction. Since objects appear to be in the direction from which the light comes, refraction, by changing the course of the light, causes an apparent displacement of the source. An example is found in the classic experiment of Kleomedes, who showed that a coin placed in the bottom of a vessel so that it is barely concealed by the sides of the latter, is apparently lifted into view when the vessel is filled with water (Fig. 483). The object at A then seems to be at the point A', and the bottom PQ of the vessel appears to be raised to $P'Q'$. Similarly, a meter rod dipped obliquely into water appears to be bent, and the divisions seem to be shortened. The latter effect is also observed when the rod is normal to the surface. This change in the apparent distance of objects seen normally through a refractive medium is to be considered as an example of refraction, although there is no deviation of the light. It will be shown in §667 that these effects are the result of differences of velocity of light in the media concerned.

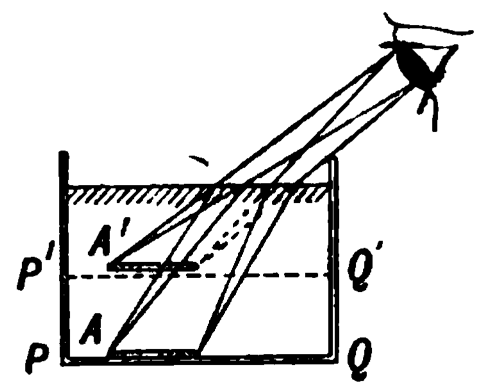


Fig. 483.

638. Intensity of Light.—The brightness of light as estimated by the eye is not capable of precise physical determination. It depends to a large extent upon the color of the light and the sensitiveness of the eye. The only consistent way in which intensity of radiation may be determined or expressed is in terms of energy.

If radiation travels through a homogeneous medium in straight

lines, and if the medium is perfectly transparent and does not itself emit radiation, the same total amount of energy must flow per second through any spherical surface concentric with the source. It follows that the *intensity* or quantity of energy passing through unit area per second, must vary inversely as the square of the distance from the source (§259).

The above conclusion is based upon the assumption that the radiation diverges uniformly in straight lines in all directions. It is not true if the medium is of varying refractivity, on account of partial reflection and of changing divergence of a cone of light in passing from one medium to another. In case a beam is made parallel by a lens or mirror there is no change of intensity with distance except that due to absorption or to imperfect parallelism.

639. Photometry.—The eye can form no exact estimate of degrees of intensity, but it can determine with great accuracy whether two adjacent surfaces are equally illuminated by lights of the same color. Upon this principle are based the different methods of **Photometry** or comparison of intensities. Two of the simplest and oldest types of photometer are the Rumford shadow and the Bunsen grease spot photometers. In the use of both it is assumed that the light from the two sources compared contains the different colors in the same proportions, making comparison possible.

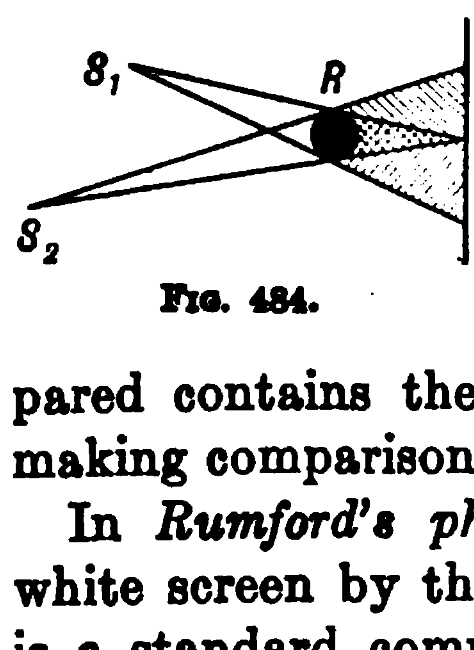


FIG. 484.

In *Rumford's photometer* shadows of a rod R are cast on a white screen by the sources S_1 and S_2 (Fig. 484), one of which is a standard comparison source. By adjusting the positions and distances of S_1 and S_2 the shadows may be made to touch and to be of equal intensity. When this is the case, it is evident that the intensity of light from each source is the same at the screen, since each shadow is illuminated solely by the source which casts the other shadow. If this intensity is I, and if the intensities of the sources at unit distance are respectively I_1 and I_2,

$$I=\frac{I_1}{r_1^2}=\frac{I_2}{r_2^2}$$

or the respective intensities of the two sources are directly as the

squares of their distances from the screen when the latter is equally illuminated by both. This relation holds likewise in the use of the other forms of photometer described below.

The *Bunsen photometer* consists essentially of a grease spot on a screen of white paper. Such a spot is more translucent than the clean paper, and for this reason appears darker by reflected light (since there is less light reflected from the spot). If such a screen is placed between sources which equally illuminate it with light of the same quality (same proportions of different colors) the grease spot will disappear. The loss in light reflected from the spot on one side will then be compensated by the increased amount transmitted from the other side.

The *Joly diffusion photometer* consists of two rectangular blocks of paraffin separated by a piece of tin foil. Paraffin is a translucent substance which appears to scatter light throughout its entire mass. If this photometer is placed between two sources of light with the tin foil at right angles to the line joining them each block will be illuminated by one source alone. If the intensity of illumination is the same on both sides the boundary line between the two blocks will disappear; if it is not the same, the boundary is clearly seen, the block receiving the smaller amount of light appearing darker than the other throughout its entire mass.

640. Lambert's Law. A flat flame or an incandescent sheet of metal appears to be equally bright whether viewed normally or obliquely to its surface. The intensity or the energy falling per second on unit area of the total surface BC (Fig. 485) is equal to the total energy emitted per second from AB at the angle α with the normal to the surface, divided by BC, or if E_α is the emissivity of AB per unit area in that direction, $E = E_\alpha (AB/BC)$. The normal emissivity is E_n and observation shows that $E = E_n$. Therefore

$$E_n = E = E_\alpha (AB/BC) \text{ or } E_\alpha = E_n \cos \alpha$$

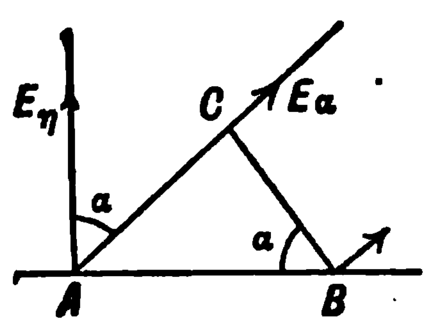


FIG. 485.

This is known as Lambert's law. In accordance with this principle, an incandescent sphere when viewed from a distance appears to be a uniformly illuminated disk.

The law does not apply to a surface bounded by an absorbing atmosphere, which will of course exercise greater total absorption in an oblique than in a normal direction. The sun, for example, which is surrounded by an absorbing atmosphere of gases, appears (as clearly shown in photographs) to be darker at the edges than at the center.

In the same way it may be shown that if I_n is the intensity of light falling normally on a screen, the illumination, when the light is incident at the angle i is

$$I_i = I_n \cos i$$

VELOCITY OF LIGHT.

641. Velocity of light.—The sensation of light is produced by a disturbance originating in distant bodies, and it may naturally be assumed that this disturbance travels with a finite velocity. Galileo, about 1600, appears to have been the first to attempt to measure this velocity. His method was substantially the same as that ordinarily used to determine the velocity of sound in the atmosphere.

Two observers stationed at some distance from each other endeavored to note the instants at which flashes of light from one station were observed at the other. The failure of such attempts made it clear that the velocity of light is so great that the time required to pass over ordinary distances is too small to be measured except by methods much more refined than those at that time available. It was natural, therefore, that the first results should have been obtained by astronomical methods, in which the distances employed are those between heavenly bodies.

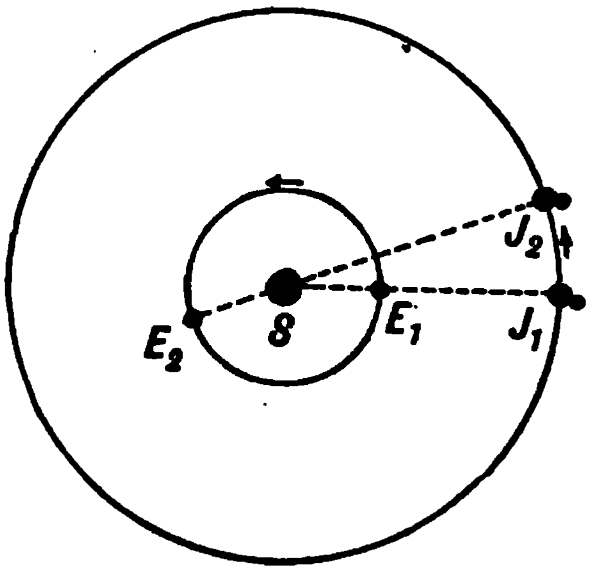


FIG. 486.

In 1675 Römer, a Danish astronomer, observed that the eclipses of Jupiter's satellites by that planet recur at regularly increasing or decreasing intervals, according to the earth's position with respect to Jupiter. If the first observations are made when Jupiter and the earth are on the same side of the sun and in line with it, the interval between the first and the second eclipse of one satellite is about 1 day 18.5 hours, but as the earth proceeds in its orbit the interval between eclipses slowly increases, so that when the earth is on the opposite side of the sun from Jupiter, the eclipse occurs about 16 minutes later than the time calculated from the first observed interval.

Römer explained this as being due to the finite velocity of light. The distance between the earth and Jupiter having in the interval increased by the diameter of the earth's orbit, the last installment of light that comes from the satellite before eclipse has this additional distance to travel and in consequence reaches the earth later by 16m. 41.6s (according to modern

observations). This and the best determinations of the diameter of the earth's orbit give 298,300 kilometers per second as the velocity of light.

642. Bradley's Method.—Römer's explanation was discredited until long after his death, when an entirely different astronomical method confirmed his views. In 1727 Bradley, the astronomer royal of England, discovered an apparent negative parallax of the fixed stars; that is, an apparent displacement not opposite to the direction in which the earth was moving in its orbit, but in the same direction. Bradley was for a time greatly perplexed by this phenomenon, but the chance observation of the direction of a wind vane on a boat sailing on the Thames, this direction not being that of the wind, but of the resultant of that of the actual wind and that of the virtual wind due to the motion of the boat, suggested to him that the apparent motion of the light coming from the stars might be the resultant of the actual motion of the light and its relative motion with respect to the moving earth. If a stone S (Fig. 487) is dropped into a vertical tube T which is at the same time moving parallel to itself in a direction at right angles to the path of the stone, the latter will have a horizontal component of relative motion with respect to the tube and will strike its side. When the tube reaches T_1, the stone reaches S_1 and the relative path of S with respect to T is the dotted line PS_1. Similarly a beam of light which actually moves with a finite velocity parallel to the axis of a telescope tube will strike the side of the latter on account of its displacement due to the motion of the earth. If the apparent angular displacement is α, it is evident that $\tan \alpha = u/V$, where u is the component velocity of the earth at right angles to the line of sight and V the velocity of light. Bradley gave the name *aberration* to this apparent angular displacement of the light from the stars. The best determinations of α, the aberration constant, is 20.445″, which, combined with the known velocity of the earth in its orbit, gives a value for V of 299,920 kilometers per second

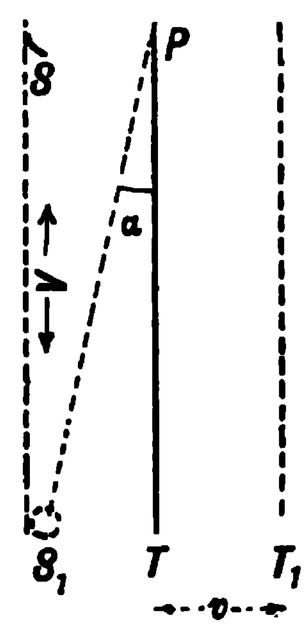


FIG. 487.

643. Fizeau's Method.—The first to make a direct determination of the velocity of light was Fizeau, who in 1849 found the time required for light to pass between Suresnes and Montmartre, near Paris, a distance of 8633 meters. His method was as follows: Light from a source S (Fig. 488) is reflected from a piece of plate glass m, focused by a lens L on the circumference F of a toothed wheel W, and, after passing between the teeth of the wheel, is made parallel by a second lens L_1. From this point the beam travels to the distant lens L_2, which focuses it on a mirror M. From this point the beam retraces its path to the source; but a portion of it will pass through the plate glass m to the eye E, by which it may be observed. If the toothed wheel is rapidly rotated a detached train of light waves will pass through as an

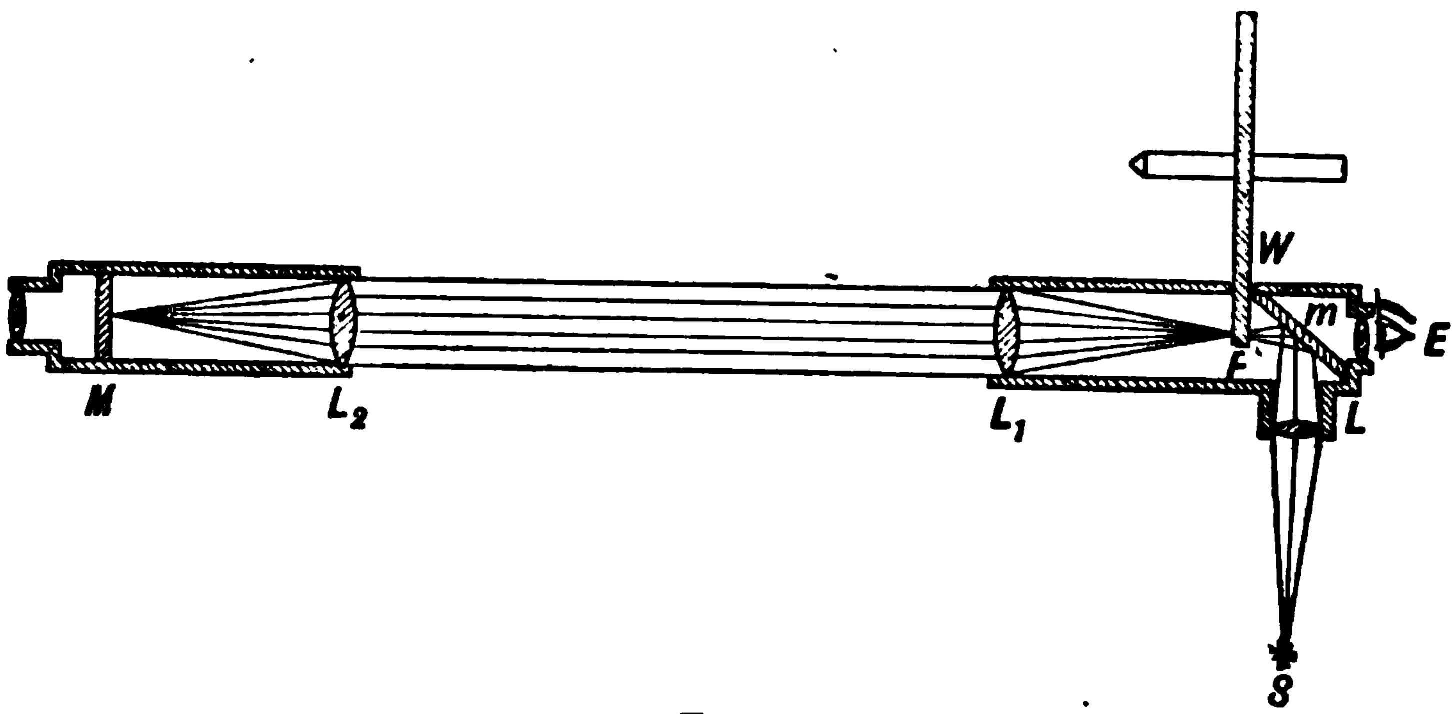


FIG. 488.

opening passes F, travel to M, and return. If in the meantime a tooth has moved into the position F the light will be eclipsed; at twice the speed required for the first eclipse the light will again reach F when an opening is at the point, and will pass to the eyepiece. At three times the original speed of the wheel the second eclipse will occur, and so on. At speeds permitting transmission of the light the waves will pass and return through the successive openings in intermittent groups, but the light will appear continuous to the eye because of the persistence of vision. From the distance between the wheel and the distant mirror and the rate of revolution of the wheel the velocity of light can be calculated.

The value of V found by Fizeau was 313,300 km./sec. Cornu, using the same method, obtained a mean result of 299,950 km./sec. from several series of experiments.

644. Method of Foucault, Michelson, Newcomb.—In 1862 Foucault determined V by means of the displacement of a beam of light reflected from a revolving mirror. The method was improved by Michelson, who made a series of observations in 1879 at the United States Naval Academy, and another in 1882

in Cleveland. Michelson's arrangement is indicated in Fig. 489. Light from a narrow slit S falls on the mirror m and is reflected to a lens L, which throws it in a parallel beam to the plane mirror M. The beam retraces its path, and if the mirror m is at rest is brought to a focus at S. If, however, m has rotated through the angle α while the light is passing from m to M and back, the reflected pencil will be rotated through the angle 2α and will form an image of the source at S_1. If the distance between S and $S_1 = d$, that between S and $m = r$, that between m and $M = L$, if n be the number of revolutions of m per second, and T the time required for light to pass from m to M and back,

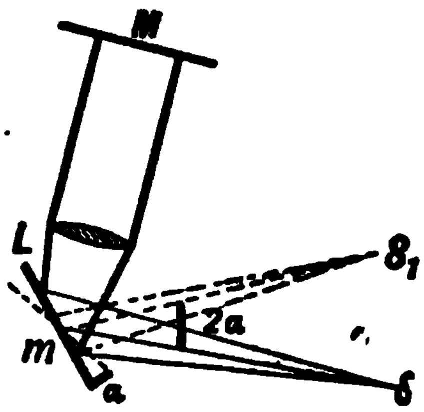


FIG. 489.

$$2\alpha = \frac{d}{r}$$

$$T = \frac{\alpha}{2\pi n}$$

$$V = \frac{2L}{T} = \frac{4\pi L n}{\alpha} = \frac{8\pi L n r}{d}$$

Foucault used a short-focus lens between S and m instead of a long-focus lens between m and M, as in Michelson's arrangement; consequently L was a short distance, not exceeding 20 meters, and the displacement d was only 0.7 mm., even when the mirror revolved 800 times per second. The result obtained by Foucault was 298,000 kilometers per second. In Michelson's experiments a long-focus lens enabled him to make r large and at the same time to throw a parallel or nearly parallel beam on M, so that the distance L could be increased indefinitely without any considerable loss of light. With a value of $L = 625$ meters, $r = 9$ m., and a speed of 257 revolutions per second, the displacement d was 133 mm. The result of Michelson's latest experiments in 1882 was $V = 299{,}850$ km./sec.

Newcomb, in 1882, made some further improvements in Foucault's method. The distance L was 3,721 meters, between the Washington monument and Fort Myer, in Virginia. The value of V obtained by him was 299,860 km./sec. The final results of Michelson and of Newcomb are probably not in error by more than 30 km./sec.

645. The Velocity of Light in Different Media, such as water and carbon bisulphide, was determined by Foucault and Fizeau, and also by Michelson; the method of Foucault being used in each case. A long tube filled with the liquid was placed between

the mirrors m and M. Michelson found the velocity in air to be 1.33 times greater than that in water, and 1.76 times greater than that in carbon bisulphide. This has an important bearing upon the choice between the emission and the undulatory theories of light (§§647, 667).

The velocity of light from all sources seems to be the same, not being appreciably affected by their intensity. Rōmer and Bradley used sunlight or starlight, Fizeau and Cornu calcium light, Foucault, Michelson, and Newcomb sunlight, Young and Forbes electric light. In space lights of all colors travel with the same velocity. This is shown by the eclipse of a white star by the moon; the star would appear red just before eclipse and blue just after if blue light travels faster then red; but no change of color is observed. It is also shown by the fact that in Michelson's experiment the light was not drawn out in a spectrum. Photographs of the spectrum of the variable star Algol, the light from which has a period of variation of about 69 hours, show that the intensities of the extreme violet and extreme red rise and fall simultaneously, proving that there is no relative retardation between them. In some material media the velocity of light of different colors differs considerably. Michelson found the velocity of blue light in carbon bisulphide to be 1.4 per cent. less than that of red. In gases this difference is inappreciable.

Light reaches the earth from the moon in about one second and from the sun in about 8.25 minutes. A small parallax has been found in the case of some o' the nearer stars, which enables rough estimates of their distances to be made. Light from one of the nearest stars, α Centauri, would require about 3.75 years to reach the earth, and that from Sirius about 17 years. It seems quite possible that a distant star may have been destroyed by an explosion or collision generations ago, and yet be visible to us by light emitted before its destruction and still on its way through space. The changes frequently observed in variable stars must take place years before they are evident to us.

THE NATURE OF LIGHT.

646. Mode of Transmission.—According to some of the older hypotheses, such as that of Descartes, light is the effect of a pressure instantaneously transmitted through a universal medium. The fact that the disturbance producing light has a

finite velocity shows, however, that it is due to motion, not to a static pressure. The radiation from such bodies as the sun heats substances on which it falls, and may produce chemical changes or electrical effects, which shows that a continuous stream of energy flows from luminous sources. According to our experience, there are only two ways in which energy may be transferred—by the actual projection of material bodies through space or by the transmission of vibrations or pulses through a stationary medium, as illustrated by different types of wave motion. Consequently there have been two rival theories regarding the propagation of light, the *emission* theory and the *undulatory* or *wave* theory.

647. Emission Theory.—Sir Isaac Newton believed that light is due to the emission of luminous particles ("corpuscles") from the source. He appears to have adopted this hypothesis chiefly because it explained the rectilinear propagation of light, for which the wave theory seemed inadequate. Newton showed by prismatic analysis that white light is a combination of many different colors. He attributed difference of color to difference in size of the corpuscles exciting luminosity.

Newton observed that water waves pass around obstacles without sensible disturbance, casting no shadows, and that sound shadows arise only under exceptional circumstances. Reasoning by analogy he could not see why light, if due to wave motion, should not travel around corners instead of in straight lines. He noticed, however, that sound waves had a greater tendency than water waves to cast shadows, and if he had carefully observed the behavior of small waves, such as ripples on water, his objections to the wave theory would probably have been removed. While large water waves pass around a pile or other comparatively small obstacle, ripples are effectually stopped, passing the object on each side without reuniting; there is a well-defined region of no disturbance, or shadow. Similarly, sounds of high pitch, due to very short waves, cast well-defined shadows.

The emission theory satisfactorily explains reflection if we suppose the corpuscles to behave like elastic spheres. If such a sphere strikes a reflecting surface at an angle i with the normal (Fig. 490) the tangential component v of its velocity will not be changed. If the magnitude of the reflected com-

ponent u is unaltered, it follows that the angle r of reflection is equal to the angle i of incidence.

Refraction is also explained if we assume that matter attracts these particles. They will then be subject to a normal acceleration as they approach the boundary, while the tangential component of velocity is unchanged (Fig. 491). If the medium offers no resistance to the motion of the corpuscles (that is, if it is transparent) it follows that the increased velocity should be maintained after entering the second medium, and that

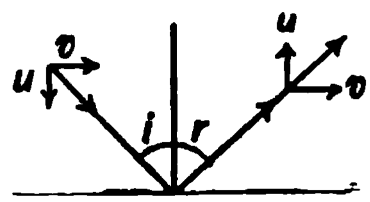


FIG. 490.

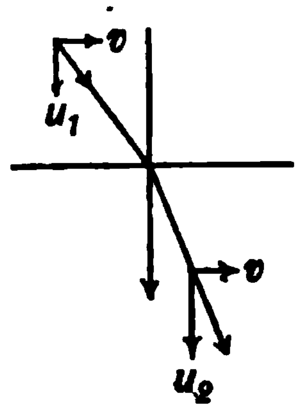


FIG. 491.

the velocity of light should be greater in more refractive media than air than it is in the latter. Experiments show that the opposite is true in all cases tested (§645). This is one grave objection to the emission theory. Furthermore, if matter attracts light corpuscles, it would be difficult to account for the enormous expulsive forces required to project the particles from luminous sources. We should also expect the speed of the particles to vary with the nature and activity of the source; and yet the velocity of light from a candle appears to be the same as that from the sun.

648. Wave Theory of Light.—Huyghens distinctly formulated this theory about 1678. He believed that space is filled with a rare medium, the ether, through which the waves are propagated from luminous bodies. This theory accounts without any difficulty for the ordinary phenomena of reflection and refraction, but was not acceptable to Newton for the reason above stated. For more than a century after Newton's time little progress was made in the subject of light, until. in 1802, Thomas Young published a paper "On the Theory of Light and Colors." In this he discussed optical phenomena from the standpoint of the wave theory, and first called attention to the fact, overlooked by Huyghens and other advocates of the wave theory, that the effect at any point of space through which light waves are passing is the resultant of the effects of a number of coincident individual awves. The magnitude of this resultant depends not only on the amplitudes, but also on the relative phases of the component

waves. If two waves of equal amplitude and moving in the same direction are in the same phase the displacement at any point is the sum of the individual displacements, and the energy, which is proportional to the square of the amplitude, is four times as great as in a single wave. If the waves are opposite in phase, the resultant amplitude and energy at any point are zero. This effect Young called the *interference* of light waves.

Young devised a simple experiment which may be regarded as a crucial test of the wave theory. Light diverging from the slit S (Fig. 492), which acts as a primary source, passes through two narrow slits S_1 and S_2 very close together, which act as secondary sources. If a screen be placed beyond these slits a series of colored and dark bands parallel to the slits will be observed on it. If one of the slits is covered the bands disappear. This shows that they are the resultant effect of two superimposed pencils of light alternately reënforcing and destructively interfering with each other. This is analogous to the interference of mercury ripples described in §258.

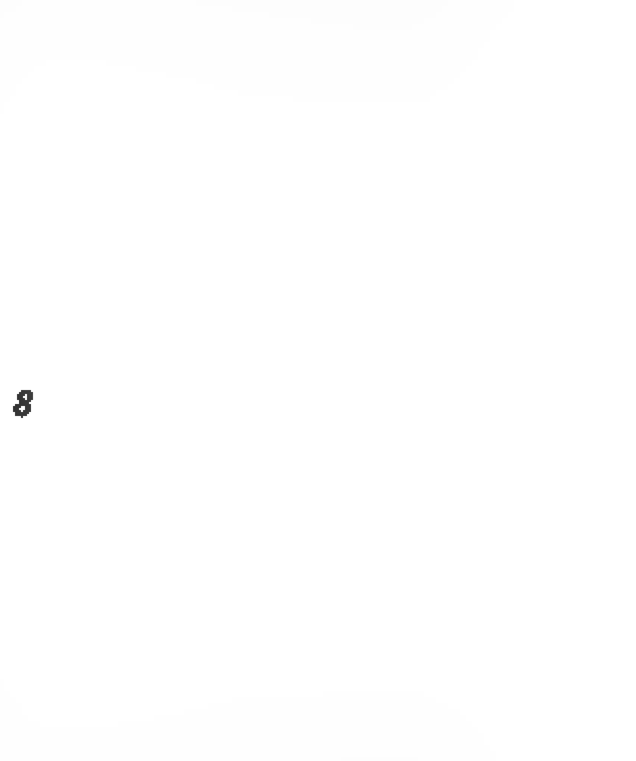


FIG. 492.

It is easy to repeat Young's experiment by ruling two narrow slits very close together on a developed photographic plate and looking through these slits at a distant electric light. The explanation is as follows: Through the slits S, S_1, and S_2 the wave disturbance propagates itself in all directions beyond the respective screens in semi-cylindrical waves having these slits as axes, as may be seen by holding a white screen in front of such a narrow slit on which light falls. It will be seen that the transmitted light diverges very considerably from the axis of the pencil, the amount of divergence increasing as the slit is narrowed (§705). There are, consequently, when two slits are used, two sets of semi-

cylindrical light waves diverging from these slits and crossing each other, as shown in Fig. 492. Along SP_0 every point of which is equidistant from S_1 and S_2, waves of all lengths from the two sources will always meet in the same phase, and there will be a maximum of white light on the screen at P_0. Along the dotted line ending at P_1 the distances of any given point from the sources differ by half a wave-length; there is destructive interference along this line and a minimum for the corresponding color at P_1. Along the line ending at P_2 the difference between the distances of any given point from the sources is a whole wave-length, so that along this line waves of the same length meet in the same phase and there is a maximum for the corresponding color at P_2. At any point P_n for which $S_1P_n - S_2P_n = n\lambda$ (n being any whole number) there will be a maximum; where $S_1P_n - S_2P_n = \frac{1}{2}(2n+1)\lambda$ there will be a minimum.

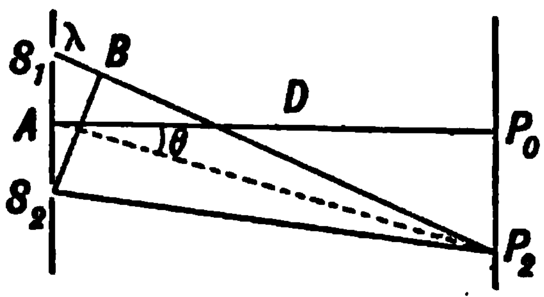


FIG. 493.

Let P_2B be equal to P_2S_2. Then $S_1B = \lambda$. Denote S_1S_2 by a. Since a is very small, S_2B is very nearly perpendicular to S_1P_2. Hence

$$\lambda = a \sin \theta$$

From this λ can be deduced. Measurements show that it is very small, being about 0.000065 cm. for red light and 0.00004 cm. for violet light.

If AP_0 be denoted by D, since AP_0 and AP_2 are very nearly equal, $\sin \theta = P_0P_2/D$. Hence $P_0P_2 = D\lambda/a$. Thus for a given distance of the screen the width of a band varies directly as the wave-length and inversely as the distance between the slits.

649. Relation between Color, Wave-length, and Frequency.—If white light falls on the slits the inner side of each band is violet, the outer side red. This shows that the wave-length is different for light of different colors, and that the wave-length of violet light is less than that of red. The central band is of course white, as all colors have a maximum at this point, regardless of their wave-length. From the relation $n\lambda = V$ (§246) where n is the frequency of vibration, λ the wave-length, and V the velocity of light, it is evident that when V changes either n or λ or both must change. If Young's experiment be performed in a medium such

as water, it is found that the width of the bands in water is to their width in air as the velocity of light in water is to that in air. Hence $\lambda_1/\lambda = V_1/V$, and n is constant. It is a matter of common experience that the color of a beam of light does not change when it enters water, hence frequency rather than wave length determines color. Color is, therefore, analogous to pitch in sound.

650. The Ether.—To account for the transmission of waves through space containing no ordinary matter it seems necessary to assume the existence of a universal medium filling all space and even interpenetrating matter itself, as shown by the existence of transparent substances. That this medium can react on matter is shown by the fact that radiant energy is transmitted from ether to matter in the case of absorption, and from matter to ether in the case of emission of radiation by material sources.

In recent years doubt as to the necessity for assuming the existence of an ether has been expressed by some who believe that it is sufficient to attribute the power of transmitting radiation to space itself. It may be doubted whether this is more than a dispute about terms. We cannot discuss the question here, but pending the settlement of the controversy it seems wise to continue the use of the word ether as at least denoting the power of space, vacant or occupied by matter, to transmit radiation.

651. Huyghens' Principle.—Huyghens assumed that a wave is propagated by every point of the medium in a wave front acting as a new center of disturbance as has already been explained and illustrated in the case of water waves (§256 and Fig. 166f). The resulting wave front is the enveloping tangent plane to the wavelets starting from these centers, as shown in Fig. 494.

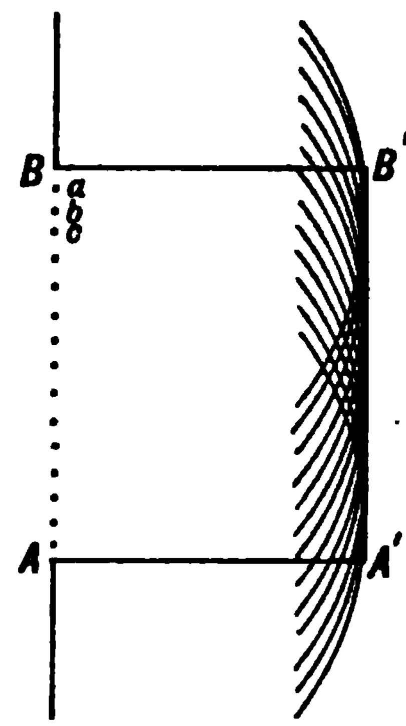


FIG. 494.

The points a, b, c, etc., between A and B (Fig. 484) taken as close together as we please, act as centers of disturbance. Along the tangent plane $A'B'$ the different waves are all in the same phase, and each point in this new tangent plane becomes a new center of disturbance, so that the resultant wave travels forward as rapidly as the disturbance is propagated from point to point of the medium.

The waves move forward without hindrance, because there is no existing displacement to oppose them; they do not travel backward, because there

is a force due to the existing displacement on the side from which the waves come sufficient to nullify the backward component of the displacement due to each successive center of disturbance. It is like the propagation of a shove through a line of people, or of elastic spheres of the same mass and elasticity; that in front is not braced to withstand the impulse, while the reaction on the one communicating the impact is expended in overcoming its forward momentum.

652. Origin and Properties of Light Waves.—Sources of light are usually bodies of high temperatures. According to the mechanical theory of heat, high temperature corresponds to a violent agitation or vibration of the ultimate particles (molecules, atoms or electrons) (§159) of matter. We may imagine that these particles impart their motion to the surrounding ether in much the same way that a tuning fork generates sound vibrations in air.

So far no evidence has been presented to show whether these waves are longitudinal, like those of sound, transverse, like those in a stretched wire, or of a more complex character, like water waves. In §753 it will be shown that the displacements in these waves must be transverse to the direction of propagation.

We may now, as a working hypothesis, assume that light is due to transverse periodic displacements in a universal medium, set up by the agitation of the ultimate particles of matter, that these waves are of different lengths (periods of vibration), but are all very short; that different colors correspond to different rates of vibration; and that waves of all lengths travel with the same velocity in free space, but with different velocities in matter. All experimental facts are in harmony with these assumptions.

In the following pages the word *ray* will often be used as a matter of convenience, *meaning thereby merely a normal to the wave front, which indicates the direction in which the wave is moving at the point considered.* The definition applies only to isotropic media (§§163, 761).

REFLECTION.

653. Reflection from a Plane Surface.—A wave diverging from the source S (Fig. 495) falls on a plane mirror MN. If the mirror were absent, the wave would at a given instant occupy the posi-

tion $AMPNB$. With the mirror in place, each element of the original wave when it reaches the mirror becomes the center of a reflected wavelet, just as it would have contributed a wavelet from the same point to form the resultant wave MPN if the mirror were absent. If, therefore, a number of circles tangent to MPN be described about centers a, b, c, etc., on MCN they must touch both the imaginary wave MPN and the reflected wave MQN. The arcs MPN and MQN are evidently similar and equal and have equal radii of curvature. If S_1 is the center of curvature of the reflected wave $SC=CS_1$, the line SS_1 is normal to the mirror, and S_1 is as far behind the latter as S is in front of it. If the eye is at E, any point reached by the reflected wave, the pencil of light entering the pupil will be focused on the retina. As the vertex of this cone is virtually at S_1, the image of the source will appear to be at that point. From the diagram it is evident that the angles i and r are equal, that is, *the angles of incidence and of reflection are equal.*

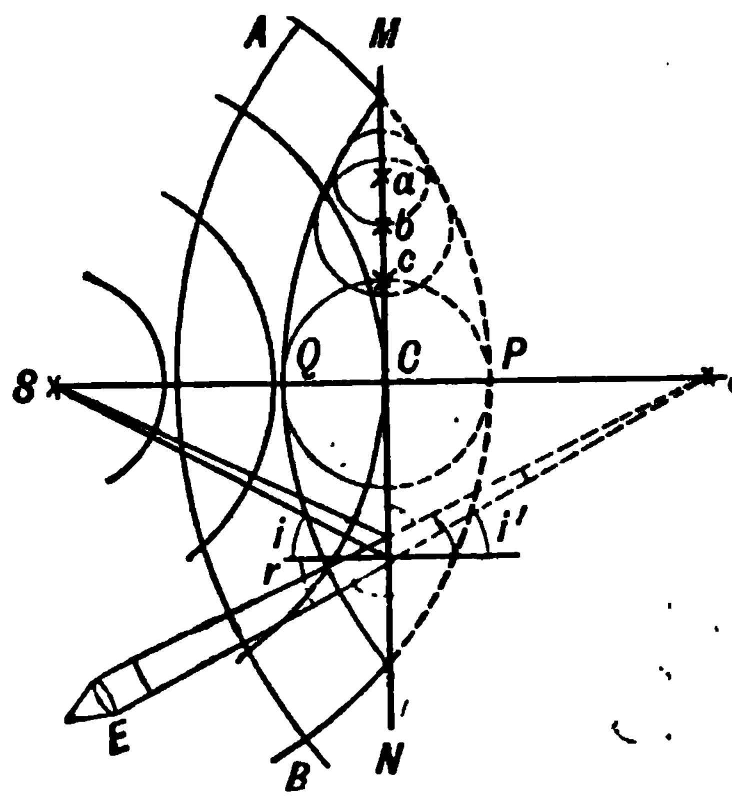


FIG. 495.

654. Focus.—The source or center of curvature of a family of waves, either divergent or convergent, is called a *focus*—literally a hearth or source of radiation. The point S from which the waves actually come is called a *real focus;* the point S_1 from which they appear to come is called a *virtual focus.* The points S and S_1 are *conjugate foci.* Since the conjugate focal distances in the case of a plane mirror are equal, it is evident that if the mirror be displaced a given distance parallel to itself the image will be displaced twice that distance.

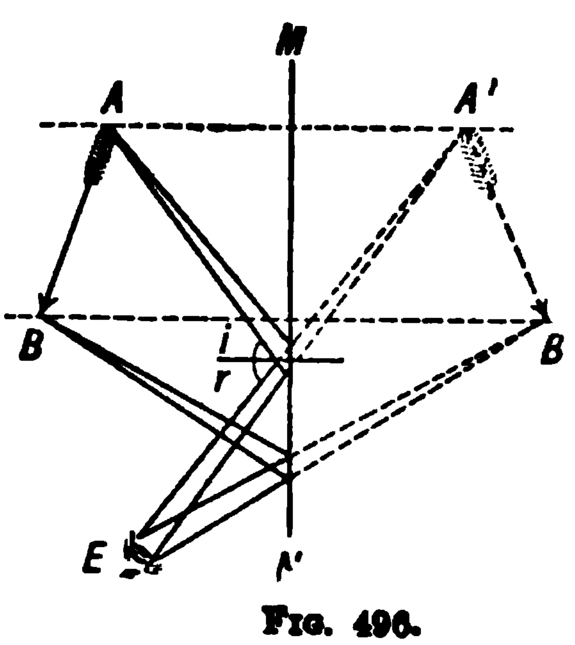


FIG. 496.

655. Images.—If $A'B'$ is the image of AB (Fig. 496), it may be shown as above that the image of each point is as far behind the mirror as the point itself is in front, and on the same normal; and that, consequently, the image and the object are symmetrically placed with respect to the mirror and are of the same size.

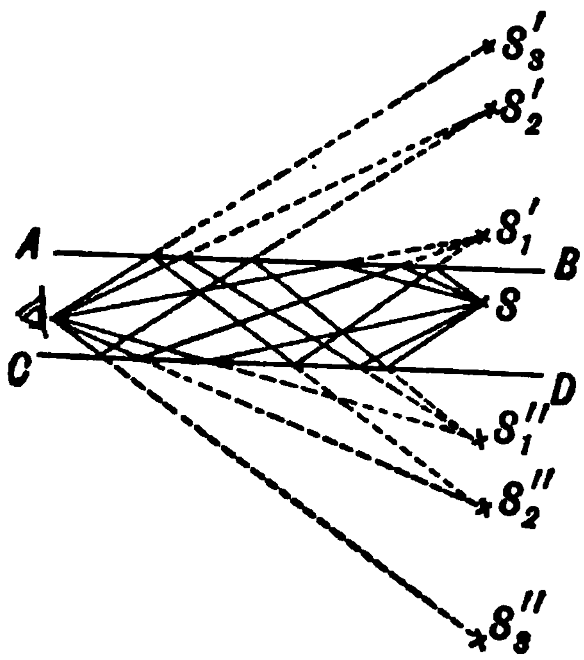


FIG. 497.

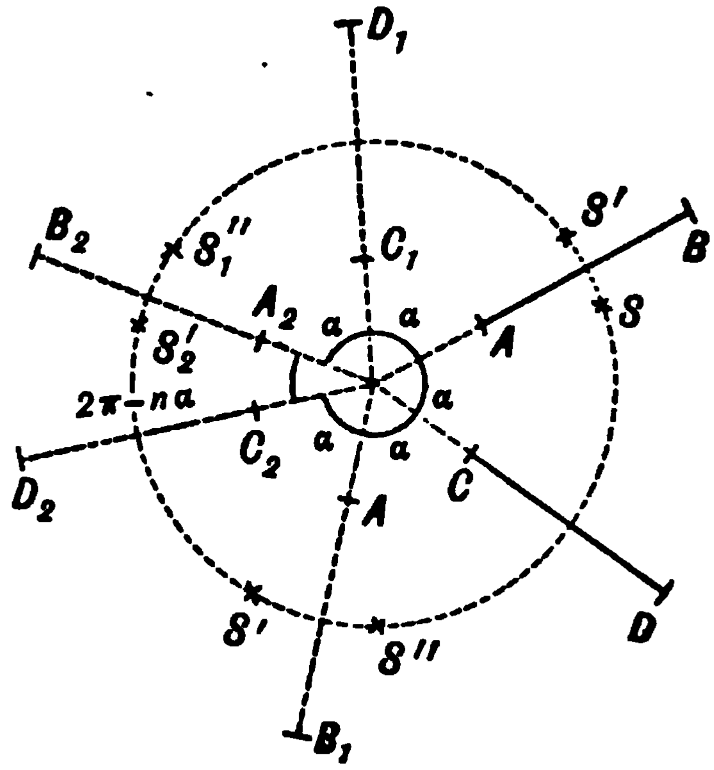


FIG. 498.

656. Multiple Reflection.—Fig. 497 shows how these images are situated in the case of multiple internal reflection from surfaces AB and CD parallel to each other. The position of these images is readily determined by the fact that the image of the first order in each surface is as far behind the surface as the source is in front, and on the same normal to the surface. The two images of the second order are fixed in the same way, by considering the images of the first order to be the sources, and so on *ad infinitum*. It is easy to see that when the mirrors are inclined at an angle α (Fig. 498) there are multiple images of the mirrors as indicated by the dotted lines, and that the successive images are symmetrically placed on each side of each mirror image and located in a circle about the point of intersection of the mirrors.

657. Reflection from Curved Surfaces.—If a wave is reflected from a curved surface the curvature of the reflected wave is changed, unless it exactly conforms to the mirror surface at incidence. Experience shows that only in a few cases is the reflected wave spherical or approximately so, and only in such cases can a definite image be formed. The ordinary type of curved mirror is that with a spherical surface. The reflected waves are approximately spherical if the diameter of the mirror is small compared with its radius of curvature. In order to determine

the position of the center of curvature and the conjugate focal relations for spherical mirrors a very simple mathematical relation is all that is required.

658. Relation between Radius of Curvature and Sagitta of Arc.—Consider the arc AB, with center of curvature C, and radius r (Fig. 499). The distance x on the bisecting radius of the arc included between the arc and the chord $AB=2y$ is called the sagitta of the arc. To determine the relation between r and x write

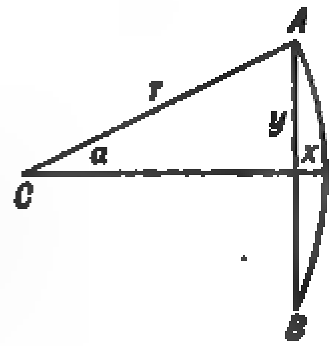


Fig. 499.

$$r^2=y^2+(r-x)^2=y^2+r^2-2rx+x^2$$

Therefore,

$$2rx-x^2=y^2$$

$$x=\frac{y^2}{2r-x}=\frac{y^2}{r(1+\cos\,\alpha)}$$

It is found that if the angle α is very small, not more than two or three degrees, the mirror will give a well-defined image. If the angular aperture 2α of the mirror is greater than four or five degrees spherical aberration becomes noticeable (§664). For all mirrors which give satisfactory images x may be neglected in comparison with r, or $\cos\,\alpha$ regarded as equal to unity, so that within the limits of errors of measurement

$$x=\frac{y^2}{2r}$$

659. Concave Mirror.—The source is at a distance u from a concave mirror MN (Fig. 500) with center of curvature at C and radius r. The waves incident on the mirror have a radius of curvature u, with a sagitta AB. Reflection begins at M and N while the vertex of the wave has still to travel the distance BD before reflection begins at D. When the vertex reaches the mirror the edges of the wave have travelled a distance $BD=AD-AB$ along MS_1 and NS_1. If the reflected wave is spherical it must have a

Fig. 500.

definite center of curvature S_1 and radius v, with sagitta DE. At the instant when reflection begins at D the incident wave, the mirror, and the reflected wave have a common point of tangency at D. If the angular aperture of the mirror is so small that the cosines of the angles α, β, and γ may be considered as equal to unity (these angles are exaggerated in the figure for the sake of clearness) we may consider the portions of the wave reflected from M and N to move parallel to the axis rather than in the directions MS_1 and NS_1. Hence

$$AD-AB=DE-AD$$
$$AB+DE=2AD$$

It is not convenient to measure sagittæ, but by using the relation developed in §658 the above expression can be transformed into one involving only the easily measured distances r, the radius of curvature of the mirror, u, that of the incident wave, and v, that of the reflected wave. The semi-chord y has the same value for all the arcs concerned, so that the common factor $y^2/2$ may be cancelled when $y^2/2r$ is substituted for AD, with similar substitutions for AB and DE. The final result is

$$\frac{1}{u}+\frac{1}{v}=\frac{2}{r}$$

The justification for the somewhat inexact assumptions made in deriving this formula is found in the fact that it agrees with experimental observations within the limits of error of measurement.

A beam of light is always reversible in direction, hence, if the source is at S_1, the image will be at S.

If the source is at a great distance from the mirror the incident wave is practically plane (parallel beam), and u is infinite. The corresponding value of v is called the *principal focal distance* f. S_1 is then the *principal focus*. The above equation then becomes

$$\frac{1}{\infty}+\frac{1}{f}=\frac{2}{r} \quad \text{or} \quad f=\frac{r}{2}$$

Hence the principal focal point is half way between the mirror and its center of curvature. The conjugate focal relation may now be written:

$$\frac{1}{u}+\frac{1}{v}=\frac{1}{f}$$

If $u > r$, $v < r$. The image is between C and the mirror.
If $u = r$, $v = r$. The image is at the center.
If $u < r$, $v > r$. The image is beyond C.
If $u = f$, $v = \infty$. The reflected light is parallel.

If $u < f$, v is a negative quantity. Fig. 501, which illustrates this case, shows that the center of curvature of the reflected wave is behind the mirror. It is a *virtual focus*, since the waves do not actually diverge from that point. It is clear that the negative sign of v indicates this result, since the distance $DS_1 = v$ is measured in a direction opposite to that in which light actually proceeds after reflection, so that the reflected light cannot pass through the point S_1.

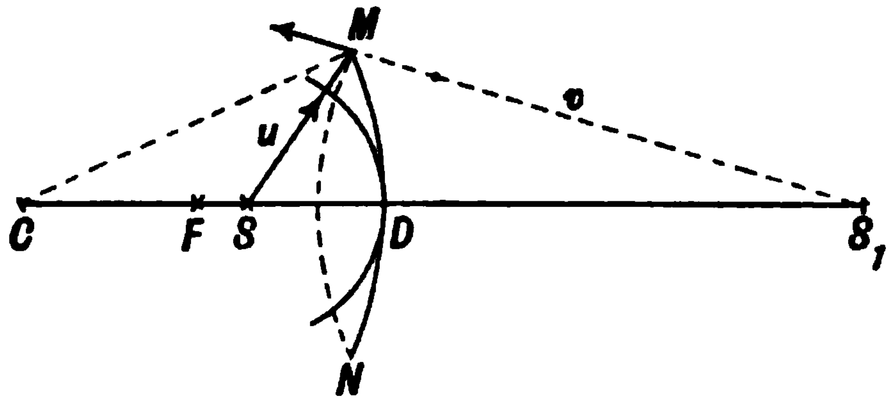


FIG. 501.

Writers differ in their conventions regarding the signs of conjugate focal distances and radii of curvature. The most easily remembered and applied as well as the most consistent rule seems to be the following:

Consider each of the quantities u, v, r as positive when it is on the same side of the mirror, as in the typical case of a concave mirror forming a real image of a real object—negative when on the opposite side.

From this rule it is evident that a positive value of v or f indicates a real focus, a negative value a virtual focus.

660. Convex Mirror.—Proceeding as in the previous case, if FG is the sagitta of the reflected wave and v its radius (Fig. 502),

$$DE + EF = FG - EF$$
$$DE - FG = -2EF$$
$$\frac{1}{u} - \frac{1}{v} = -\frac{2}{r} = -\frac{1}{f}$$

where, provisionally, v, r, and f, may be considered as mere magnitudes affected with the negative signs in the formula.

Comparing this expression with that deduced for a concave mirror, we see that it will become identically the same if we agree to consider the radius of curvature and the principal focal

distance of a convex mirror as negative in accordance with the rule given above; v is also negative since the reflected light is divergent.

The general formula applicable to all mirrors is, therefore,

$$\frac{1}{u}+\frac{1}{v}=\frac{1}{f}$$

u being usually positive, and f to be taken as positive for a concave, negative for a convex mirror. When $f=\infty$ we have the case of a plane mirror.

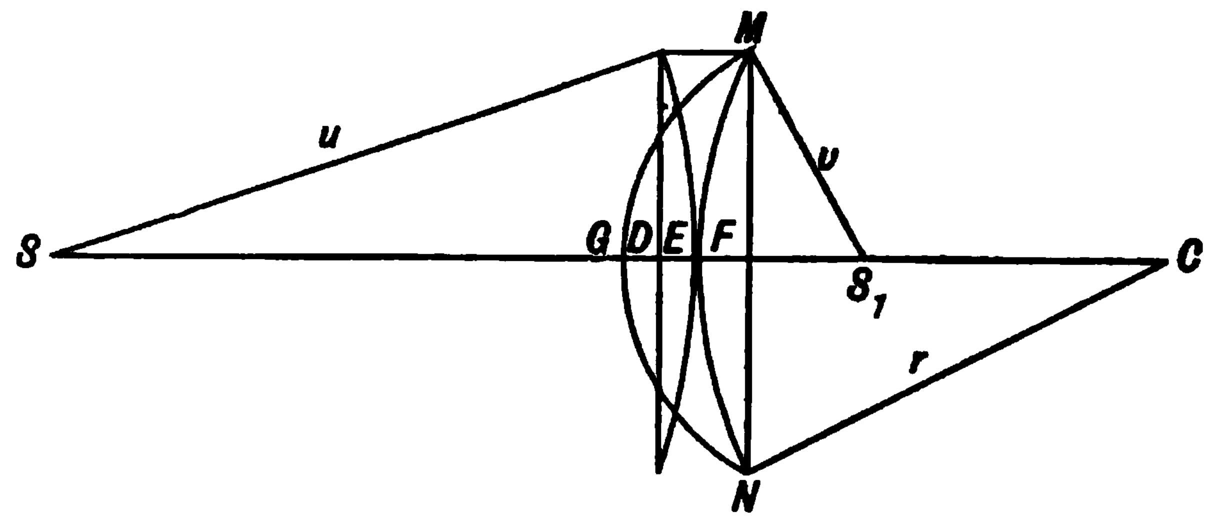


Fig. 502.

If we make f negative in the expression for v given in §659, we see that in the case of a convex mirror v is always less than f and negative. If, however, the light incident on the mirror is convergent to a point at a distance $-u$ behind the mirror, v may become positive; so that a convex mirror may give a real image of a virtual source.

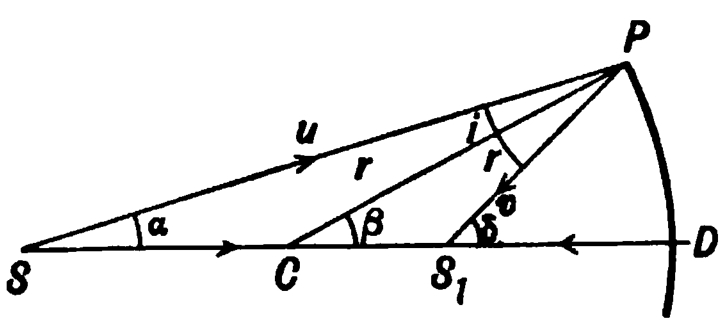


Fig. 503.

661. Geometrical Method.—The same results may be obtained by applying the law of plane reflection to "rays" without any hypothesis as to the nature of light. The ray SD (Fig. 503) will, as it is incident normally at D, be reflected back on itself. The ray SP will be reflected at P, so that the angles i and r are equal. The intersection of these two reflected rays will fix the position of the image S_1. From a well-known geometrical relation we have

$$\frac{SC}{\sin i}=\frac{SD-CD}{\sin i}=\frac{u-r}{\sin i}=\frac{u}{\sin \beta}$$

$$\frac{CS_1}{\sin r}=\frac{CD-S_1D}{\sin r}=\frac{r-v}{\sin r}=\frac{v}{\sin \beta}$$

Therefore,

$$\frac{u-r}{r-v}=\frac{u}{v}$$

From which

$$ur+vr=2uv$$

Dividing through by uvr,

$$\frac{1}{u}+\frac{1}{v}=\frac{2}{r}$$

The assumptions made in this case are that $SD=u$ and $S_1D=v$, which is a sufficient approximation to the truth when the angles α, β, and δ are small.

The formula for the conjugate focal relations of a convex mirror may be derived in the same way.

In some cases it is more convenient to use the geometrical or ray method than that of waves; but it must always be remembered that these "rays" merely represent normals to the wave front.

662. Images Formed by Spherical Mirrors.—If any two radii be drawn from any point of a source, the point of their intersection after reflection will fix the position of the corresponding

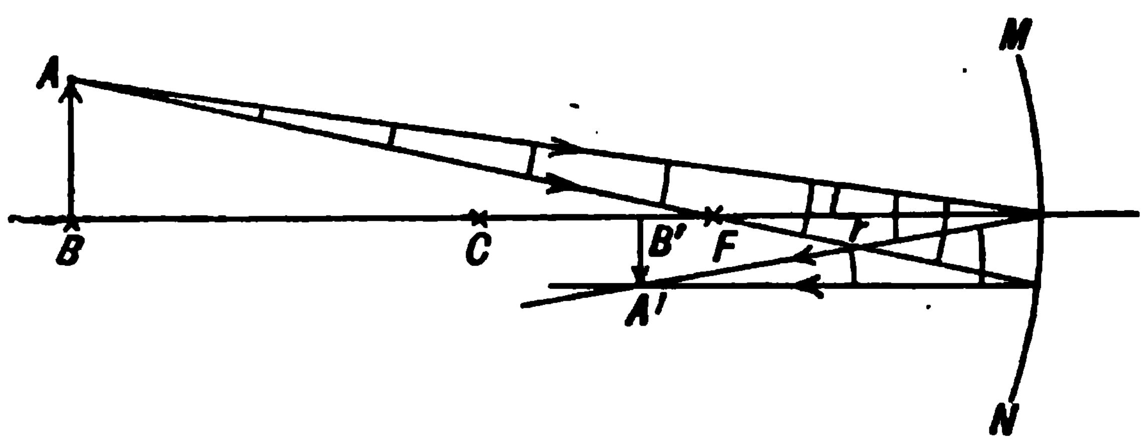


FIG. 504.

point of the image. Any pair of radii will do, but for convenience two of the following are usually chosen, because their course after reflection is easily determined: The radius parallel to the axis, which after reflection passes through the principal focus; that which passes through the principal focus, which becomes parallel to the axis after reflection; that which is incident at the intersection of the mirror with its axis.

The construction of the images formed by a concave and by a convex mirror is illustrated by Figs. 504, 505, 506, where

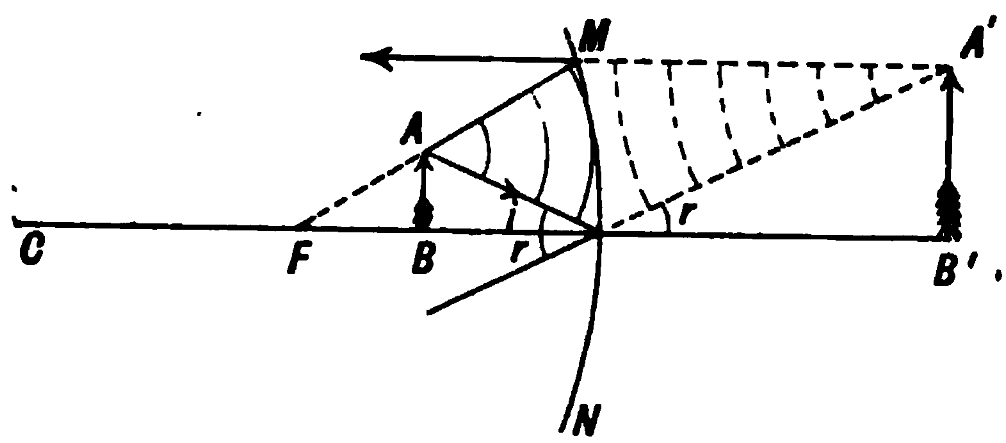


FIG. 505.

the points A' and B' are located by using the pair of rays last mentioned above. In the first, the image is real and inverted; in the second and third the images are virtual and erect.

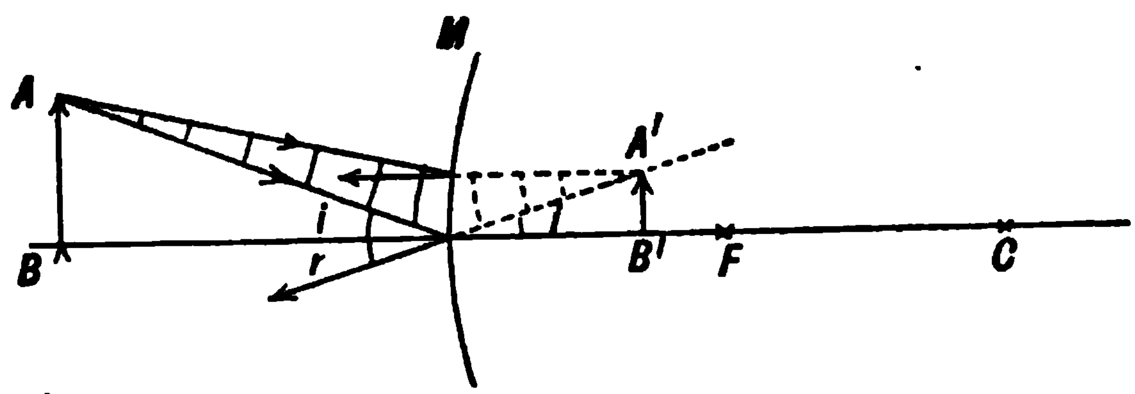


FIG. 506.

663. Magnification.—Since the angle subtended by the object at the mirror is i, while that subtended by the image is the equal angle r, it is evident that the relative sizes of the object and image are to each other as their respective distances from the mirror.

$$\therefore \quad o/i = u/v$$

The real image formed by a concave mirror may be of the same size as the object, or larger, or smaller; the virtual image is always larger, since $v > u$. The virtual image formed by a convex mirror is always smaller than the object.

664. Spherical Aberration and Caustic Curves.—If a converging wave is truly spherical there is a perfect focus at its center of curvature. As a matter of fact, the waves reflected from a spherical mirror are not perfectly spherical, except in the special case where the source is at the center of curvature of the mirror. The normals drawn from any points of the reflected wave

are tangent to the curve HF, which is called a *caustic curve.* The cusp F of this curve corresponds to the focal point of a mirror of small aperture. The light reflected from the sides of a cup containing coffee or milk plainly shows this caustic curve on the surface of the liquid.

The deviation from a spherical shape of waves reflected from a mirror of large aperture is called *spherical aberration.*

If light is obliquely incident on a mirror, the reflected waves are not spherical, even when the aperture is small, but have different radii of curvature in planes at right angles. As a result, a point image of a point source cannot be obtained, but there are two elongated images at right angles to each other and in different positions, which are called *focal lines.*

FIG. 507.

The origin of the focal lines is clearly seen if we consider the mirror MN (Fig. 508) to be part of a larger mirror MNO, on the axis SB of which the source S lies. Constructing the reflected rays incident at different points of this mirror, it is clear that, while the focal cusp of the entire mirror is at S_1, all the rays coming from MN intersect approximately at the point F_1. The diagram gives a cross-section of the incident and reflected rays. If this diagram be rotated about the axis SB by an amount equal to the diameter of the mirror MN the point F_1 will describe the arc of a circle with its center on the line SB. This is the primary focal line, which will appear on a screen placed at F_1 as a narrow curved strip. After passing F_1 all the rays reflected from MN will intersect the axis SB at various points between S_1 and F_2 (since all the planes of incidence contain SB). A screen placed at this point will show a narrow elongated patch of light, S_1F_2, the secondary focal line. If the screen is at right angles to the reflected pencil the patch of light will be approximately a lemniscate or figure 8.

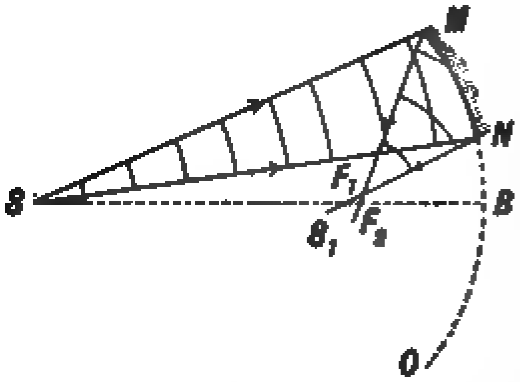


FIG. 508.

665. Cylindrical Mirror.—A parallel beam incident on such a surface is brought to a real or virtual line focus. The image of a point source is

likewise a line. Such mirrors and the reflected pencil are said to be *astigmatic*. (A pencil symmetrical about an axis, that is, having a point vertex, and thus giving a point image of a point, is said to be *homocentric.*) In the case of a concave cylindrical mirror, if the point source lies outside the principal focus, there will be a real image AB and a virtual image $A'B'$ in planes at right angles to each other, as illustrated in Fig. 509.

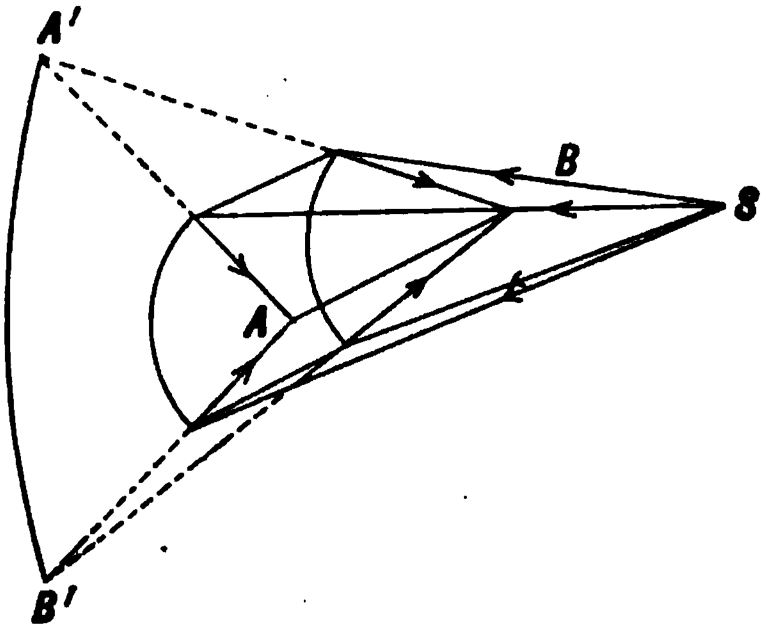


FIG. 509.

666. Paraboloidal, Ellipsoidal, and Hyperboloidal Mirrors.—The light from a point source at one focus of an ellipsoidal reflector will be brought without aberration to the other focus, a real image being formed. Light from a source at one focus of a hyperboloidal mirror will have a virtual focus at the conjugate focus of the mirror. If the source is at the focus of a paraboloidal mirror, the light will be reflected in a parallel beam; and parallel light will be brought without aberration to a real focus by such a mirror.

REFRACTION AND DISPERSION

667. The ancients were acquainted with the fact that a beam of light is more or less deviated in passing from air to water. The **Law of Refraction** was first discovered in 1621 by *Willebrod Snell.* He found by experiment that the ratio of the sines of the angles of incidence and of refraction is constant at the boundary between two media. The ratio $\sin i/\sin r$ is called n, the *index of refraction.* The angle of incidence is usually measured in air.

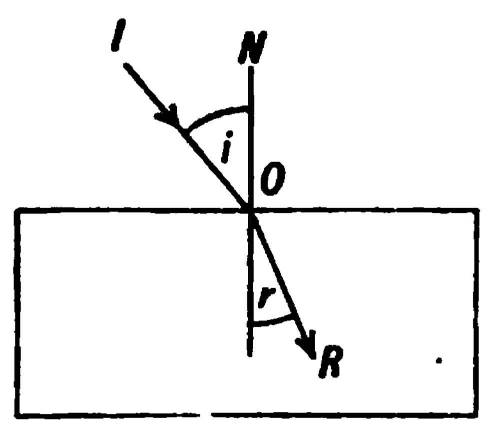


FIG. 510.

It was shown by Huyghens that refraction is very simply explained by assuming a change of velocity in passing from one medium to another. Direct measurements by Foucault, Fizeau, and Michelson show that light travels with different velocities in air, water, and carbon bisulphide (§645).

Consider a plane wave AC incident obliquely on the smooth plane surface of separation between air and another transparent medium (Fig. 511), the velocity in air being V_1 and that in the second medium V_2. A spherical wave will diverge from the point A into the second medium when the disturbance

reaches that point, and later other spherical waves successively diverge from B' and C'. While the wave travels in the first medium a distance $CC'=V_1t$ the wave from A will travel the distance $AA'=V_2t$ in the second medium. The disturbance from B will in the same time travel a distance $BB'+B'B''=(V_1+V_2)\,t/2$, if B is half way between A and C. Since $B'B''=\frac{1}{2}AA'$, a tangent plane can be drawn from C' to the two circles with centers at A and B'. It is easily shown by this method that the waves from all points in the original wave front will be tangent to the same plane, the new wave front in the second medium. Further,

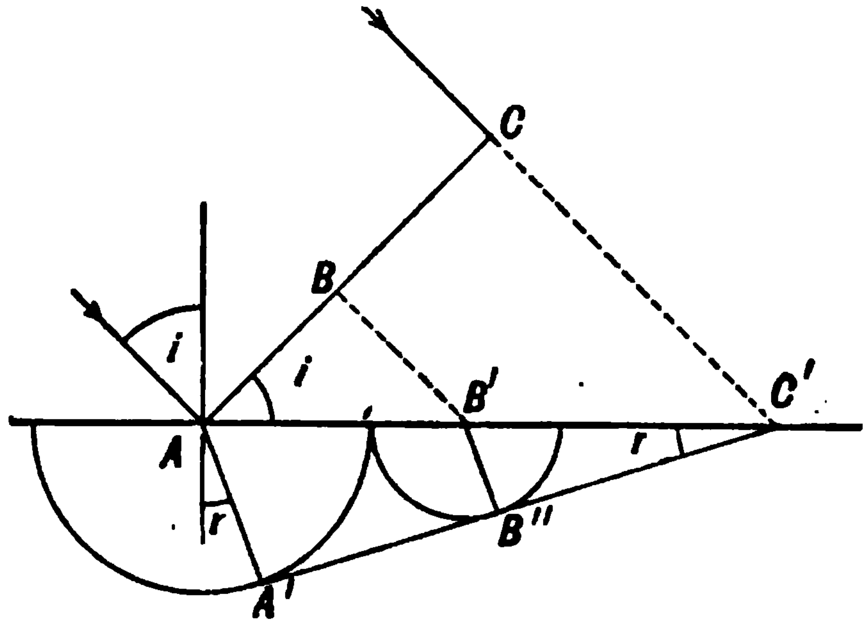


FIG. 511.

$$CC'=AC'\sin i=V_1t$$
$$AA'=AC'\sin r=V_2t$$

Therefore,

$$\frac{\sin i}{\sin r}=\frac{V_1}{V_2}=n$$

The physical significance of the constancy of the sine ratio discovered by Snell thus becomes apparent. The student should always think of the index of refraction as being the ratio of the velocities of light in the two media, rather than as the ratio of the sines of two angles. The latter mode of statement conveys no clear physical idea, and, moreover, seems to break down in the case of normal incidence.

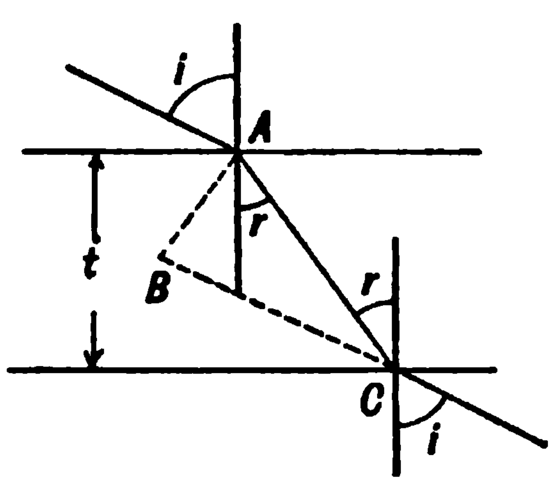


FIG. 512.

668. Medium with Parallel Surfaces.—An incident pencil will be deflected in one direction on entering the second medium of thickness t and an equal amount on reëntering the first medium, as shown in Fig. 512. The course of the pencil will then be parallel to its original direction, but there will be a lateral displacement AB.

669. Image due to Refraction at Plane Surface.—When an object is viewed normally to the boundary (Fig. 513) there is no lateral displacement, but only an apparent change in distance. Waves from an object S at a distance d below the surface of the medium travel with a velocity V_2 to the point B, where the vertex of the wave enters air, in which the velocity is V_1. The disturbance then travels a distance BB' in air while another portion of the wave still within the second medium travels the distance $AA' = V_2t$. The center of curvature of the emergent wave is at S_1, a distance d_1 below the surface. There is a virtual image of the source at this point. If the cone has only a small divergence, $AA' = BC$, the sagitta of the wave in the refracting medium, BB' is that of the wave in air, and $d = AS$ and $d_1 = AS_1$ their respective radii of curvature; hence, from the relation previously used (§658).

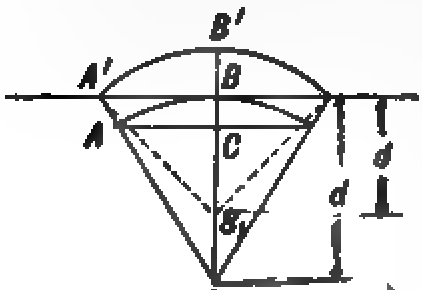


FIG. 513.

$$AA' = y^2/2d = V_2t$$
$$BB' = y^2/2d_1 = V_1t$$

Therefore,

$$\frac{d}{d_1} = \frac{V_1}{V_2} = n, \quad \text{or } d = nd_1$$

The angle of the cone of light entering the eye is limited by the size of the pupil, and is, therefore, very small, so that the use of the above method is justified. The apparent depth of the object below the surface is $d_1 = d/n$. There is an apparent displacement toward the observer amounting to $(d - d_1) = (n-1)\,d/n$. It is thus made clear why the depth of a pond appears to be less than it actually is, and why objects immersed in water appear to be shortened. Since the index of refraction is about 1.33, a pond six feet deep seems to be only about four and a half feet in depth.

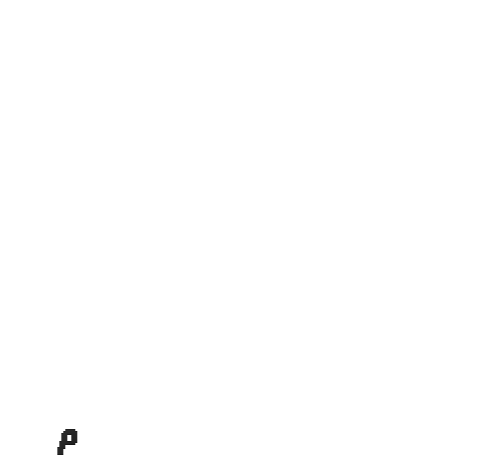


FIG. 514

If the cone is wide there is considerable aberration, as shown

in Fig. 514. This is not apparent to the eye, which limits the aperture of the effective pencil, except through a slight lateral displacement (the image being at Q if the eye is at E).

The index of refraction of plane parallel plates may be obtained from the relation deduced above. A microscope is focused on a small object on a table, such as a pencil mark O (Fig. 515). When the plate is placed over the mark it will be necessary to raise the microscope a distance d to bring the virtual image o into focus. The apparent depth of the object below the surface is $t' = t/n$, and $d = t - t' = t - t/n$. Hence

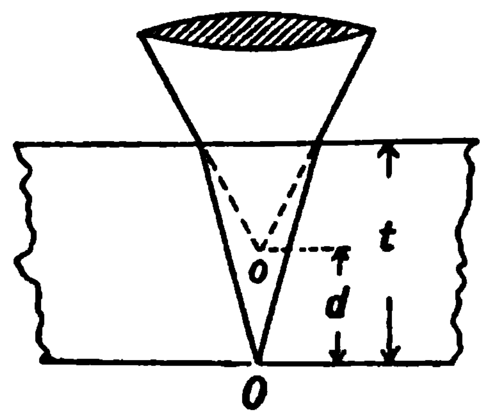


FIG. 515.

$$n = \frac{t}{t-d}$$

670. Prism.—If light waves pass through a transparent medium bounded by plane surfaces which are not parallel, the deviation of the incident pencil on entering the first surface is not exactly compensated on emerging from the second surface. If the source is at S (Fig. 516) the image, or center of curvature of the wave within the prism, is at S_1 and that of the emergent wave is at S_2. To determine the deviation of the pencil and the positions of the foci S_1 and S_2 it is convenient to follow the course of given wave normal or "ray." The intersection of pairs of such rays will fix the position of the desired foci or centers of curvature of the waves.

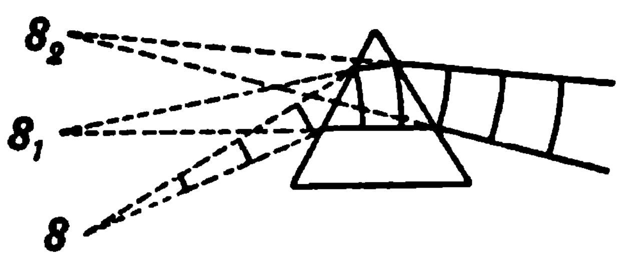


FIG. 516

671. Deviation—Minimum Deviation.—The total deviation of a given ray is $D = D_1 + D_2$ (Fig. 517).

$D_1 = i_1 - r_1$; $D_2 = i_2 - r_2$;
$D = i_1 + i_2 - (r_1 + r_2)$
But $r_1 + r_2 = A$, since
$B + A = 180° = B + r_1 + r_2$.
Therefore, $D = i_1 + i_2 - A$

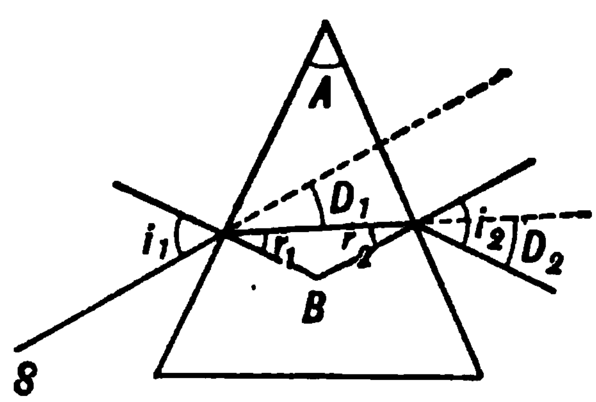


FIG. 517.

It is easily shown, experimentally or mathematically, that D

has a minimum value when $i_1 = i_2$, in which case the incident and emergent ray are symmetrical with respect to the refracting angle of the prism. In this case

$$i_1 = i_2 = \frac{D+A}{2}$$

$$r_1 = r_2 = A/2$$

Therefore,

$$n = \frac{\sin i}{\sin r} = \frac{\sin \frac{1}{2}(D+A)}{\sin \frac{1}{2} A}$$

This relation is commonly used for determining the index of refraction of substances in the prismatic form. The angles of the prism and of minimum deviation are measured with a spectrometer (§720). As the index is not the same for different colors, it is evident that the prism can be set at the angle of minimum deviation for only one color at a time.

672. Dispersion of Color.—The index of refraction of any given substance varies with the color, or wave length; consequently the deviation caused by a prism will not be the same for all colors.

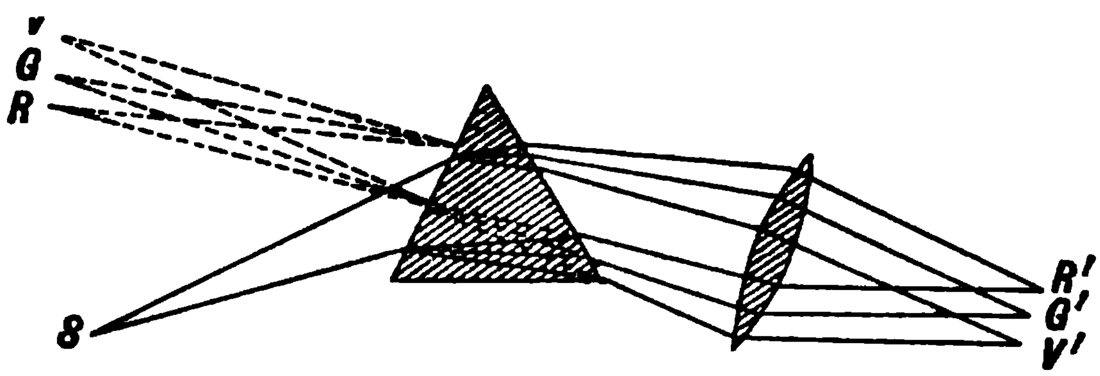


FIG. 518.

Consider a narrow source S such as an illuminated slit parallel to the edge of the prism (Fig. 518). If the source emits red light alone, a virtual red image of the slit is observed at R. If green and violet light are also emitted, a green and a violet image are seen at G and V. Real images of these colors may be formed at R', G', and V' by a lens. Such a group of slit images is called a *line spectrum*. This separation of the colors is called *dispersion*. If the source emits waves of an infinite number of lengths included between the red and the violet, the infinite number of partially overlapping images of the slit will form a *continuous spectrum*. If the slit is wide the different colors will greatly over-

lap, and the spectrum is said to be impure. There is less overlapping when the slit is narrowed; but, since no slit can be made infinitesimally narrow, it is manifestly impossible to obtain a perfectly pure spectrum.

673. Fraunhofer Lines.—If a wide slit illuminated by sunlight is used, a continuous spectrum is observed, apparently like that given by a candle flame. Such a spectrum was observed by Newton. If, however, the slit is very narrow, it will be seen that a number of fine dark lines parallel to the slit cross the spectrum. These lines were first seen by Wollaston in 1802. He observed a virtual spectrum by looking directly through a prism at an illuminated slit. Fraunhofer, about 1815, by the use of better prisms, and by forming a real image of the spectrum with a lens, was able to find several hundred of these lines, which are now usually referred to as Fraunhofer lines. It is evident from this that the solar spectrum differs from that of a candle in not being absolutely continuous. The dark gaps in the position of different colors show the absence of corresponding images of the slit, and therefore the absence of these colors in the sunlight. In the section on Absorption it will be shown that most of these dark lines are due to the absorption of light of definite wave lengths by vapors in the solar atmosphere §737).

The Fraunhofer lines may be used as reference points in measuring indices of refraction of prisms for different colors. The more prominent lines were labeled by Fraunhofer with letters of the Roman and Greek alphabets. Some of the more important of them are the A line (really a group of fine lines), due to absorption by the earth's atmosphere; the neighboring D lines, due to sodium vapor in the sun; the F line, due to hydrogen; the H and K lines, due to calcium. These lines are shown in the reproduction of the solar spectrum (upper part of Fig. 581).

674. Dispersive Power.—The deviation of a particular color by a prism increases with the index of refraction. The angular separation or dispersion between two colors depends on the difference between their respective indices of refraction. If a prism has a very small refracting angle, the angles of incidence, refraction, and emergence of a given pencil transmitted at the angle of minimum deviation will likewise be small, and the sines

of these angles may be considered as equal to the arcs; consequently,

$$n=\frac{\sin \frac{1}{2}(A+D)}{\sin \frac{1}{2}A}=\frac{A+D}{A}$$

Therefore,

$$D=(n-1)A$$

If D_1 and D_2 are the deviations of two given colors, the Fraunhofer lines C and F, for example, and D_3 that of an intermediate color halfway C to F, D_2-D_1 is the angular dispersion of the extreme colors and D_3 is the mean deviation of the spectrum of angular width D_2-D_1. The dispersive power d of the prism is the ratio of the angular dispersion of the two colors to their mean deviation, or

$$d=\frac{D_2-D_1}{D_3}=\frac{(n_2-1)A-(n_1-1)A}{(n_3-1)A}=\frac{n_2-n_1}{n_3-1}$$

Newton assumed that the ratio of dispersion between two given colors to the mean deviation, or the *dispersive power*, is the same

	n_D	n_F-n_C	$\frac{n_F-n_C}{n_D-1}$
Water	1.3330	0.0060	0.0180
CS_2	1.6303	.0345	.0547
Ether	1.3566	.0052	.0149
Alcohol	1.3597	.0062	.0174
Crown glass	1.5160	.0073	.0141
Light flint glass	1.5718	.0113	.0197
Heavy flint glass	1.7545	.0274	.0363
Very heavy flint glass	1.9625	.0488	.0507
Quartz	1.5442	.0078	.0129
Diamond	2.4173	.0254	.0179
Iodide of silver	2.1816	.1256	.1063
Air (0° C., 760 mm.)	1.00024289	.00000295	.0121
H_2	1.00014294	.00000195	.0136
CO_2	1.00044922	.00000460	.0102

for all substances, but Dollond, in 1757, showed that this is by no means the case. Two different prisms may have the same value of $n_3—1$, but very different values for n_2-n_1, or conversely.

The table shows the values of n_D and the dispersive power between the C and F lines for some substances, the mean deviation being that corresponding to the D lines. There are great differences between the refractive and dispersive powers of different specimens of glass.

675. Irrational Dispersion.—The dispersive power of a given prism (for equal increments of wave-length) varies in different parts of the spectrum, usually increasing toward the violet. There is no simple ratio between deviation and wave-length, hence such spectra are said to be *irrational*. If for any three colors the ratio $(n_3 - n_1)/(n_2 - 1)$ were the same for all substances the spectra formed by different prisms would all be alike in the distribution of colors, and one spectrum would be simply a larger or smaller copy of any other. As stated above, this ratio is not the same for different substances, so that the spectra formed by different prisms are also irrational with respect to each other. It is possible, for example, to make a prism of crown glass and one of flint glass which will give spectra of equal length between the lines A and K; but it will be found that the positions of the other Fraunhofer lines do not coincide in the two spectra, as they would if the dispersion were rational.

The following table showing the differences between the refractive indices of various substances for the A, D, F, and G Fraunhofer lines illustrates irrationality of dispersion. It will be seen that the ratio $(n_F - n_D)/(n_D - n_A)$, for example, is not the same for the different substances.

	$n_D - n_A$	$n_F - n_D$	$n_G - n_F$	$\dfrac{n_F - n_D}{n_D - n_A}$
Crown glass	0.00485	0.00515	0.00407	1.062
Heavy flint glass	.01097	.01271	.01062	1.158
Water	.00409	.00415	.00344	1.015
CS_2	.01898	.02485	.02446	1.309

Although as a general rule the index of refraction increases as wave-length diminishes, there are exceptions, as described under the head of anomalous dispersion (§777).

676. Achromatic and Direct-vision Prisms.—The unequal dispersive power of different substances is utilized for making prismatic combinations for producing deviation with very little dispersion (Fig. 519), or dispersion without deviation of the spectrum as a whole (Fig. 520). These two types are respectively called *achromatic* and *direct-vision* prisms.

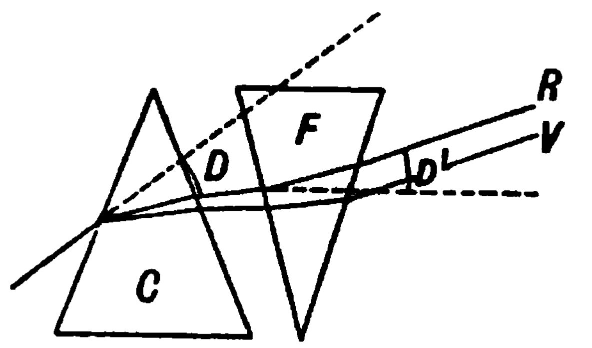


FIG 519.

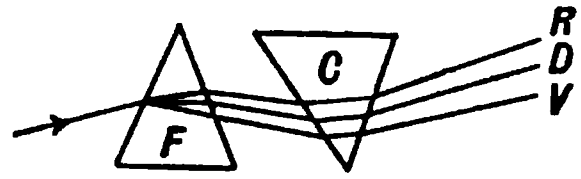


FIG. 520.

LENSES

677. Lenses are transparent bodies, generally with spherical surfaces, which form images by changing the divergence of light waves. The ordinary types of single lenses are shown in Fig. 521. The first three forms, known as double-convex, plano-convex, and concavo-convex, are *thicker at the center than at the edges*. If surrounded by a less refractive medium, the central portion of the incident wave is more retarded than the edges by these lenses, and the curvature of the wave is diminished or reversed in direction. These lenses have, therefore, a convergent effect. They are called *convex* or *converging* lenses. In the second group, embracing the double-concave, plano-concave, and convexo-concave lenses, *the edges are thicker than the center*, so that the outer portions of an incident wave are more retarded than the center. The curvature of the wave is increased and the lenses have a divergent effect. Such lenses are called *concave* or *diverging* lenses. If the two types of lenses are placed in a more refractive medium there is a reversal of these effects.

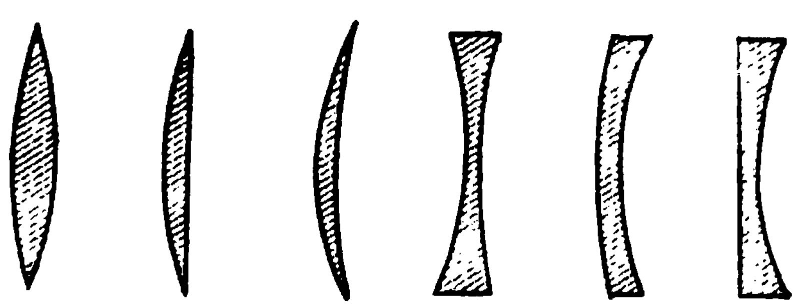


FIG. 521.

678. Equivalent Air Path or Reduced Optical Path.—At a given instant a wave front is in a given position; later it will be in a dif-

ferent position, and may have its orientation and curvature greatly modified by reflection or refraction. The one condition that must always be fulfilled, if the wave is to preserve its identity, is that the time required for the disturbance to travel from a point in the original wave front to the corresponding point in the new wave front is the same for all parts of the wave. For example, the disturbance traveling from S by the path $SPRQS_1$ (Fig. 522) reaches S_1 at the same time as the disturbance leaving S at the same instant and traveling along $SAES_1$. The latter has been sufficiently retarded by passing through a greater thickness of glass to compensate for the greater distance in air $SPRQS_1$. Similarly, the time required for the wave to travel from P to Q is the same as that from B to D. In comparing the distances traversed in equal times in different media, account must be taken of the velocity of light in the respective media. For example, in Fig. 522, $PR+RQ=V_1t$; $BD=V_2t$. Therefore, $PR+RQ=(V_1/V_2)BD=nBD$. If BD is the distance actually traversed in a medium of refractive index n, the *equivalent air path* or *reduced optical path* is nBD.

679. Conjugate Focal Relations.—Consider the case of a double-convex lens of refractive index n surrounded by air, the refractive index of which may be taken as unity. Let the radius of curvature of the first surface of the lens be r_1, that of the second r_2.

FIG. 522.

Let u be the distance of the source from the lens. PB is a section of the incident wave front of radius u, and QD that of the emergent wave front, of radius v.

The disturbance actually travels radially from P to R, thence to Q, but if α is very small, the path in air may be assumed to be

equal to $P'Q'$ without appreciable error. Placing the optical path through the center of the lens equal to this distance, we have

$$P'Q' = AB + BC + CD + DE = n(BC + CD),$$

or

$$AB + DE = (n-1)\ (BC + CD)$$

Substituting reciprocal radii of curvature for sagittæ (§658), this becomes

$$\frac{1}{u} + \frac{1}{v} = (n-1)\left(\frac{1}{r_1} + \frac{1}{r_2}\right) = \frac{1}{f}$$

In this u, v, r_1, and r_2 are considered as mere lengths, that is, numerical quantities without sign. As we shall later treat them as algebraical quantities with signs, it may be noted that, in the above case, u and v are both measured in the direction in which the light proceeds. It should also be noticed that the refraction at the first surface makes the wave less divergent, that is, it tends to converge it toward the opposite side. The same is true of the second surface. Hence both surfaces may be described as converging surfaces. If the curvature of either surface were opposite to its direction in the double-convex lens, it would be a diverging surface.

If the source of light is at an infinite distance, that is, if the incident waves are plane, $u = \infty$ and $v = f$. Hence f is the distance of the point called *the principal focus*, to which the lens converges plane waves.

In the case of a double concave lens, of thickness CD along the axis (Fig. 523), if the incident wave front is PB and the emergent wave front QF, put the optical path BF equal to the optical path PQ (assumed parallel to the axis, since α is small).

$$AC - AB + n(CD) + DE + EF = n(AC + CD + DE)$$
$$\therefore AB - EF = (n-1)\ (-AC - DE)$$

Substituting radii of curvature for sagittæ,

$$\frac{1}{u} - \frac{1}{v} = (n-1)\left(-\frac{1}{r_1} - \frac{1}{r_2}\right) = -\frac{1}{f}$$

The above formula would be identical with that for the double convex lens if the signs of v, r_1, and r_2 were reversed. Now

each of these quantities, in the case of the double concave lens, is actually measured in the opposite to the direction in which it is measured in the double convex lens. Hence, if we take the latter direction as positive for each, and so treat all of them (and v) as algebraical quantities, the two formulæ will become identical.

From Fig. 523 it is evident that incident light is made more divergent by the lens. In fact both surfaces are diverging surfaces. The significance of the negative sign of f in the above formula is that the principal focus is virtual, its distance from the lens being measured in a direction opposite to that in which the light actually travels.

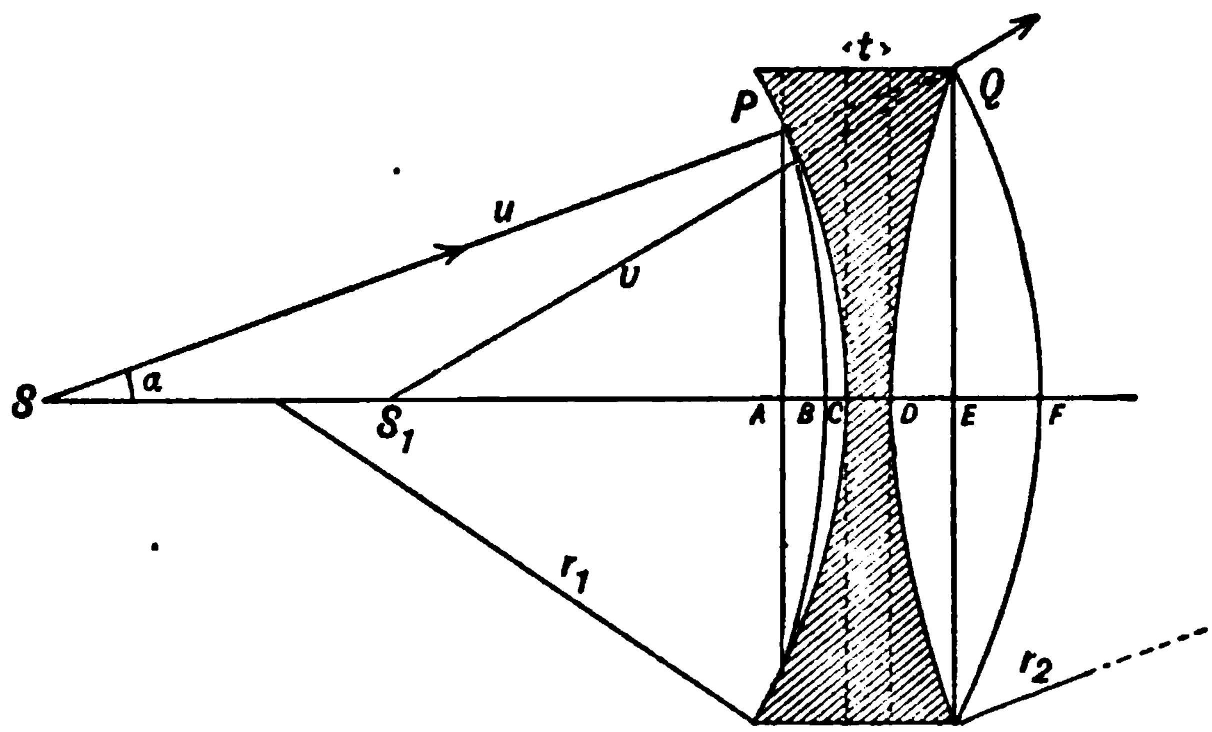


FIG. 523.

By applying the same method to the other types of spherical lenses it will be found that the general solution of all cases is the formula

$$\frac{1}{u}+\frac{1}{v}=(n-1)\left(\frac{1}{r_1}+\frac{1}{r_2}\right)=\frac{1}{f}$$

provided we adopt the following rule regarding signs:

Consider each of the quantities, u, v, r_1, r_2, f, as positive when it is on the same side of the lens as in the typical case of a double convex lens forming a real image—negative when on the opposite side.

Negative values of v and f indicate that the light diverges from

a virtual focus after passing through the lens. These conventions are consistent with those of (§659).

The following cases arise when f is positive:

When $u=\infty$, $v=f$, the principal focal distance.

When $u > f$, v is positive and there is a real conjugate focus.

When $u=f$, $v=\infty$. The transmitted beam is parallel.

When $u<f$, v is negative and greater than u for all positive values of u, and there is a virtual conjugate focus.

680. Axes of Lens.—The line passing through the centers of curvature of the surfaces of a lens is called the ***principal axis***. In every lens there is a point on the principal axis, called the *optical center*, which has the property that no ray passing through it is deviated in direction, although there is more or less displacement, depending on the thickness of the lens.

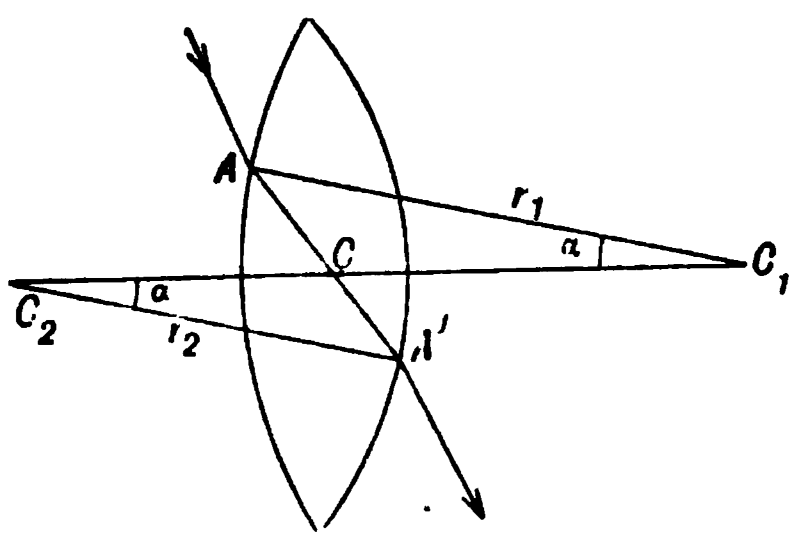


FIG. 524.

The existence of this point may be shown thus: Let two parallel radii of curvature r_1 and r_2 (Fig. 524) be drawn to the two surfaces of a lens. Since the two plane elements of the lens A and A' are parallel, being respectively perpendicular to two parallel lines, the refracted ray AA' is propagated in a medium with parallel sides and emerges parallel to its original direction. Since the triangles ACC_1 and $A'CC_2$ are similar,

$$\frac{r_1}{CC_1}=\frac{r_2}{CC_2}$$

This is true whatever may be the value of the angle α, therefore C is a fixed point, the optical center of the lens. All ray paths which pass through this point are called ***secondary axes***. In the case of a thin lens, the center of the lens and the optical center may usually be regarded as coincident.

681. Images by Lens.—The image of A (Fig. 525, a, b, c) must lie on the secondary axis AA', that of B on the secondary axis BB'. Rays drawn parallel to the principal axis from the points A and B pass through the principal focus F and intersect the lines

AA' and BB' at the points A' and B', which determine the position and magnitude of the image. Since the point A' lies above the principal axis when the image is on the same side of the lens as the object, and below it when the image is on the other side of the lens, it is evident that all virtual images formed by a single lens are erect, all real images inverted.

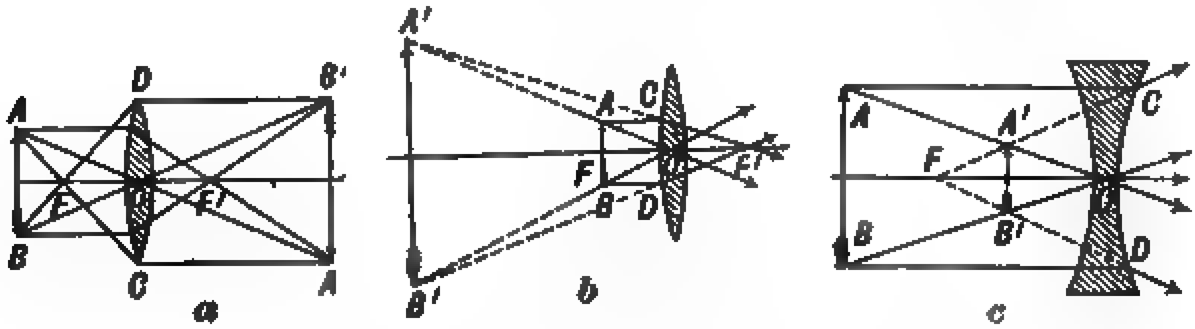


Fig. 525.

Since an object AB and its image $A'B'$ subtend equal angles from the center of the lens (the angle included between the secondary axes AA' and BB') it is evident that their relative sizes are proportional to their respective distances from the lens, or

$$\frac{AB}{A'B'} = \frac{u}{v}$$

682. Spherical Aberration.—In deriving the formula for the conjugate focal relations of lenses it has been tacitly assumed that the emergent wave is spherical. With lenses of small aperture this is shown by experience to be practically true; but when the aperture becomes large there is noticeable spherical aberration. This is illustrated by Fig. 526.

Fig. 526.

While the central part of the wave travels from B to C the edge of the wave will travel along LMN the distance $LMNO = PQ = nBC > LMN$. It is evident, therefore, that the edge of the emergent wave (represented by the dotted curve) will pass through O instead of N, and will

have a greater curvature toward the axis than if the wave were spherical, with S_1 as a center. The rays, instead of converging to S_1, as shown in the lower half of Fig. 526, will cross each other as shown in the upper half, being enveloped by a caustic surface instead of by a right cone.

683. Correction of Spherical Aberration.—If the rays passing through the edge of a lens are stopped by a diaphragm which permits only the central portion of the incident pencil to pass the spherical aberration will be greatly reduced. It is also possible to grind surfaces slightly differing from a spherical form, so that for a given pair of conjugate focal distances the emergent wave is truly spherical. Such lenses are called *aplanatic*. In some cases when the conjugate focal distances differ greatly, spherical aberration may be reduced by making the two surfaces of the lens of different curvatures. Consider, for example, a

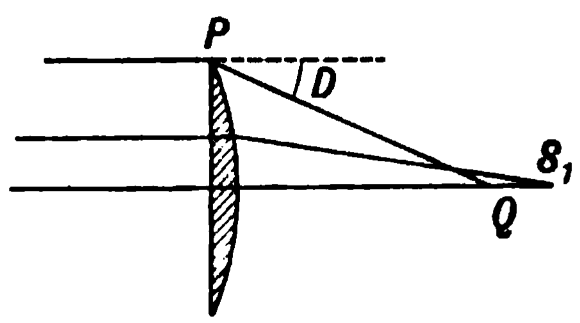


Fig. 527.

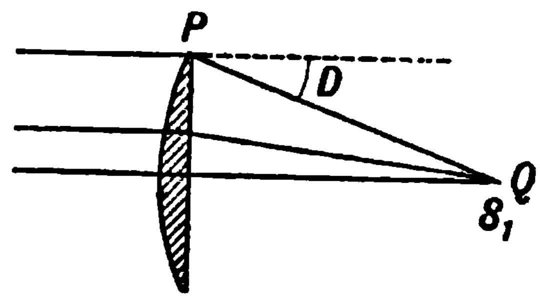


Fig. 528.

plano-convex lens of great aperture (Figs. 527, 528) first with the plane then with the convex face toward a source so distant that the incident light is parallel or nearly so. If we consider the deviation of the ray PQ in each case, it is evident, on recalling the condition for minimum deviation by a prism, that in the second case the angle D will be less than in the first, because the refracting edge of the lens is then more nearly in the position with respect to the incident and emergent rays which gives minimum deviation, and consequently the nearest approach of the ray PQ to the focus S_1.

One form of thick lens of great angular aperture commonly used as part of microscopic objectives is almost entirely free from spherical aberration.

Suppose that a ray of light which starts from S in a transparent sphere of radius R is refracted at O along a line that intersects CS produced in S_1. We shall show that, if $CS = R/n$, then $CS_1 = nR$, and therefore all rays from S seem, after refraction, to come from one point S_1.

Using a well-known geometrical principle we may write

$$\frac{CS}{R}=\frac{\sin r}{\sin \alpha};\ \frac{CS_1}{R}=\frac{\sin i}{\sin \beta}$$

$$\therefore \frac{\sin r}{\sin \alpha}=\frac{1}{n}=\frac{\sin r}{\sin i}$$

$$\therefore\ \alpha=i$$

Since the angle C is common to the two triangles COS_1, COS

$$\beta+i=\alpha+r$$

$$\therefore\ \beta=r$$

Hence

$$CS_1=R\frac{\sin i}{\sin \beta}=R\frac{\sin i}{\sin r}=nR$$

This is an exact relation, no matter how large the angle i may be, so that an object in the lens at S would have a virtual image at S_1 entirely free from aberration. The same is practically true if the segment of the sphere below S is removed and the object placed in contact with the surface. In practice this lens is often a hemisphere, the object being placed at such a distance below the plane side that its virtual image formed by refraction at the plane surface corresponds in position to the point S. There is some aberration in this case due to refraction at the lower surface. In the method of "oil immersion" the object is placed at S and the space between it and the hemispheral lens is filled in with an oil of the same refractive index as the lens.

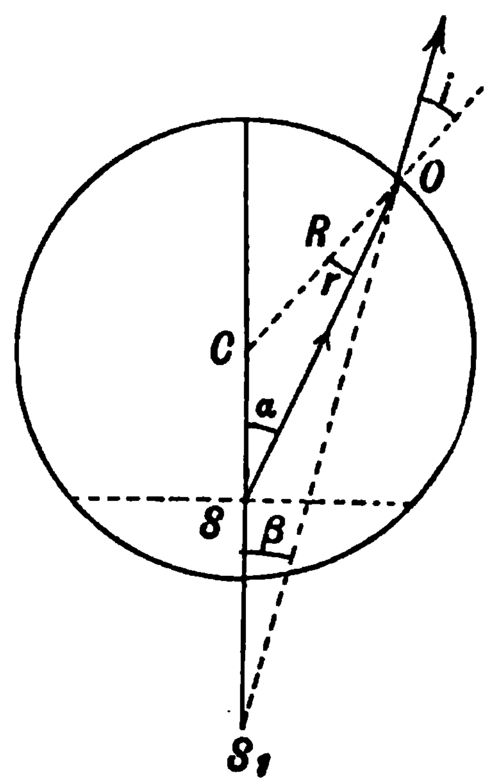


Fig. 529.

684. Focal Lines.—If a pencil of light falls obliquely on a converging lens, instead of a point image two real focal lines will be formed, like those due to a concave mirror. If the lens is divergent, these focal lines will be virtual. The formation of these lines by a converging surface is made clear by considering the effect of a single refracting surface PQ, imagining it to be extended to R, so that SS_1 is the principal axis (Fig. 530). The rays transmitted through the actual refracting surface PQ, will, by reason of spherical aberration, pass through a narrow arc through F_1 with its center on SS_1. This is the *primary focal line*. These rays will all intersect the axis SS_1 between S_1 and T. The normal cross-section of this pencil is a narrow lemniscate-shaped region at F_2, the *secondary focal line*, at right angles to the primary focal line. The second refarcting surface of the lens will modify but not change the general character of this result

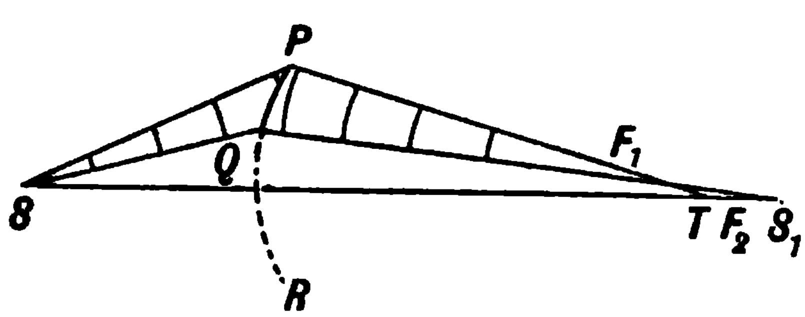


Fig. 530.

685. Cylindrical Lens.—The effect of such a lens is like that of a cylindrical mirror. A point source S has two linear imgess, as shown in Fig. 531, one AB parallel and the other $A'B'$ at right angles to the axis of the lens. The image AB parallel to the axis is at a distance given by the relation

$$\frac{1}{u}+\frac{1}{v}=(n-1)\left(\frac{1}{r_1}+\frac{1}{r_2}\right)$$

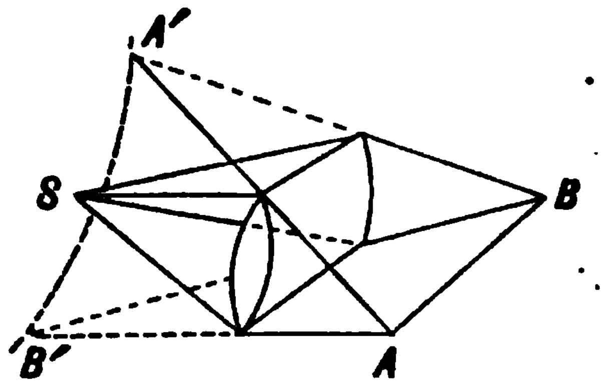


FIG. 531.

and may be either real or virtual; the other, $A'B'$ is virtual and may be considered as due to each longitudinal strip of the lens acting as a prism of the same angle. Any lens with different curvatures in planes at right angles to each other will give similar focal lines or astigmatic images.

686. Combinations of Lenses.—If the lenses are thin, with principal focal lengths f_1 and f_2, and so close that the distance between them may be neglected,

$$\frac{1}{u}+\frac{1}{w}=\frac{1}{f_1};\ -\frac{1}{w}+\frac{1}{v}=\frac{1}{f_2}$$

If w, the focal distance conjugate to u, is positive with respect to the first lens, it is negative with respect to the second. Therefore,

$$\frac{1}{u}+\frac{1}{v}=\frac{1}{f_1}+\frac{1}{f_2}=\frac{1}{f}$$

This expression is generally true for either converging or diverging lenses if the proper signs are given to f_1 and f_2.

687. Chromatic Aberration.—Since

$$\frac{1}{f}=(n-1)\left(\frac{1}{r_1}+\frac{1}{r_2}\right)=(n-1)K$$

it is evident that the principal focal distances are different for different colors, being less for violet than for red (Fig. 532). There is no way to remedy this defect in a single lens, but it may be greatly reduced by a suitable combination of lenses.

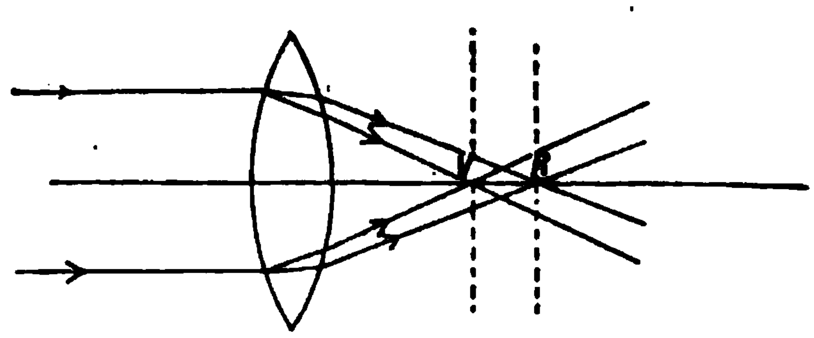


FIG. 532.

688. Achromatic Combinations.—By combining two or more lenses of different dispersive powers, two or more given colors

may be brought to the same focus, just as prisms may be combined to give deviation without dispersion (Fig. 533). If two lenses are used, for each color

$$\frac{1}{f_1}+\frac{1}{f_2}=\frac{1}{f}$$

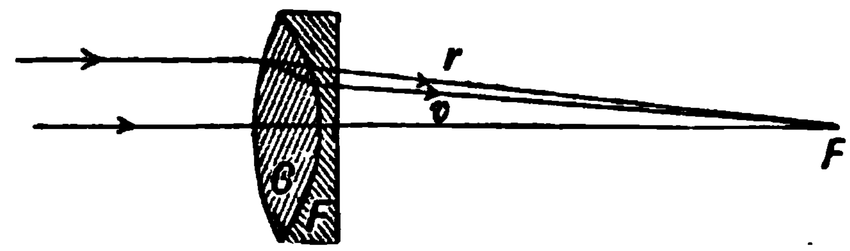


FIG. 533.

if the lenses are in contact. If we wish to combine the two colors corresponding to the C and F lines, f must be the same for both.

If n_F' and n_C' be the refractive indices of the first lens, n_F'' and n_C'' those of the second,

$$\frac{1}{f}=\frac{1}{f_F'}+\frac{1}{f_F''}=(n_F'-1)K_1+(n_F''-1)K_2$$

$$=\frac{1}{f_C'}+\frac{1}{f_C''}=(n_C'-1)K_1+(n_C''-1)K_2$$

therefore,

$$(n_F'-n_C')K_1=(n_C''-n_F'')K_2$$

The values of $K_1=1/r_1'+1/r_2'$ and $K_2=1/r_1''+1/r_2''$ may be arbitrarily chosen to satisfy this relation. Since $n_F>n_C$ it is evident that K_1 and K_2 must be of opposite sign, so that either f_1 or f_2 must be negative If f_2 is negative and greater than f_1, f is positive and the lens is convergent. If f_2 is negative and less than f_1 the combination is divergent. Usually the positive lens is of crown glass, the negative of flint, and they are shaped to fit close together, so that $r_2'=r_1''$ and often $r_2''=\infty$ (Fig. 533).

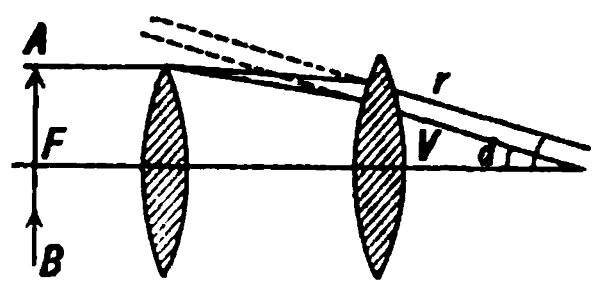


FIG. 534.

Chromatic aberration may also be reduced by using two lenses of the same index of refraction at a certain distance d from each other. To take a specific case, if the second lens is placed at a distance from the first equal to its own focal length, the rays of different colors which diverge from each other at the first lens will be made approximately parallel by the second. If an object is placed at the principal focal point F (Fig. 534) of the combination, a virtual image at infinity will be formed, and, as shown by the figure, the violet and the red images will subtend approximately equal angles α at the eye, and will, therefore, be superimposed on the retina.

REFRACTION PHENOMENA

689. Total Reflection.—If a ray of light travels from a more to a less refractive medium, the angle of emergence i is greater than

the angle of incidence (which, being in the more refractive medium, may still be called r for consistency). Since $\sin r = (\sin i)/n$, and since i has a maximum limit of 90 degrees, r has a maximum limit k such that $\sin k = 1/n$. No pencil incident on the boundary at a greater angle than k, the *critical angle*, can emerge. It will, therefore, be totally reflected (CC', Fig. 535). Since $\sin k$ varies inversely as the index of refraction, the critical angle is different for different colors. Violet will first be subject to total reflection as r increases, and finally the red.

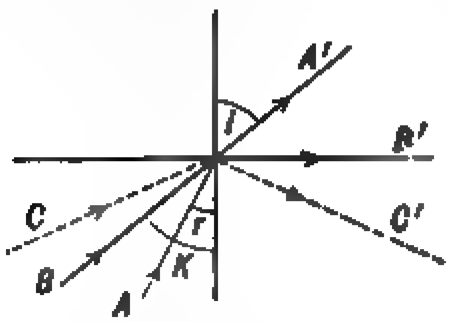


FIG. 535.

FIG. 536.

A parallel-sided plate cannot be used to show total reflection, since any pencil entering such a plate must emerge at the same angle. Objects of prismatic form are best adapted for the purpose. The effect may be seen by looking down through the side of a glass containing water or at a test-tube sunk in water. A fish can see objects throughout the space above the water, but he sees them through the limited cone of angle $2k = 97°$ (Fig. 536), arranged around a circle, with tops pointing inward.

Some values of k are given below:

Water	48° 36′	Quartz	40° 22′
Crown glass	43° 2′	Diamond	24° 26′
Flint glass	37° 34′		

The smaller the critical angle of a jewel with regular facets, the greater the proportion of light totally reflected by it. This explains the great brilliancy of the diamond.

The index of refraction of a liquid or of a small portion of an opaque object may be determined by measuring the angle of total reflection from its surface when in contact with a more refractive medium and using the relation $\sin k = n/n_1$, where n is the index of the less and n_1 that of the more refractive medium.

690. Transition Layer.—It seems quite possible that the change of index of refraction at the boundary is not abrupt, but that there is a transition layer t due to interpenetration of the two media, or occlusion at the surface causing a gradual change in the index. If this be the case, total reflection may be considered as altogether due to refraction. When the angle of incidence is equal to or greater than k the wave front in the transition layer will swing around and become normal to the surface (Fig. 537); then the lower edge will gain on the upper and the wave will swing back into the first medium. If we consider an air film between two refracting media the two transition layers may encroach on each other (Fig. 538), in which case the lower edge of the wave will be retarded, and a part of it will pass into the third medium. It might be expected, therefore, that if

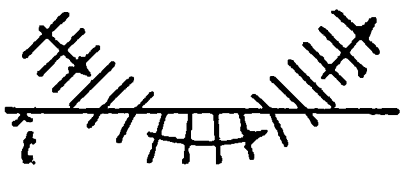


Fig. 537.

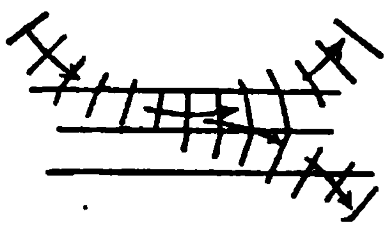


Fig. 538.

the air film from which total reflection takes place is very thin total reflection will cease. This has been found to be the case. If a right-angled total reflecting prism with a slightly convex hypothenuse surface is pressed against a glass plate, total reflection takes place from the hypothenuse of the prism when the angle of incidence i is sufficiently large, but some light will always be transmitted through the region surrounding the point of contact even where the air film has a measurable thickness. It is found that the thickness of the air film through which transmission can occur (which may be considered as approximately the thickness of the transition layer) differs with the wave-length and with the angle of incidence, and may reach several thousandths of a millimeter.

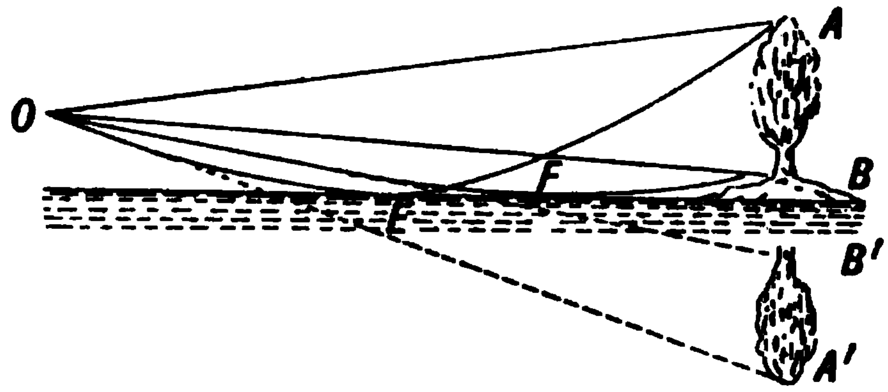


Fig. 539.

691. Mirage.—Examples of the type of total reflection referred to above are found in the case of refraction by gases of varying density. This phenomenon is called mirage. The air above a furnace or a heated surface such as a pavement exposed to the sun's rays rapidly increases in density and refractive power in going upward. If the line of vision forms a small angle with

the surface, distant objects are seen apparently reflected from the surface. The formation of one type of mirage is shown in Fig. 539. The object AB is viewed directly through the pencils OA, OB, while an inverted image $A'B'$ is also seen, due to the refraction of the pencils OEA, OFB, by the heated air near the ground. This is one of several types of atmospheric mirage. Other types showing distortion or displacement of objects are due to local differences of temperature in the atmosphere, which cause changes in density and refractive power. They are very easily seen by viewing objects at a grazing angle across heated surfaces. Similar effects are to be seen by looking through sheets of glass with irregular surfaces, or non-homogeneous mixtures of liquids, such as water with an excess of salt crystals at the bottom of the vessel, or with alcohol above and imperfectly mixed with it.

When the sun is near the horizon, the rays reaching the eye traverse strata of air of gradually increasing density, which cause them to bend downward. For this reason the sun is visible when it is actually below the horizon a distance about equal to its own diameter.

The scintillation of stars is due to a similar cause. Their apparent direction and intensity are subject to rapid fluctuations as masses of air of varying density drift across the line of sight.

If sunlight be focused on a small hole beyond which it diverges to a screen, a jet of coal gas, hydrogen, or carbon dioxide will cast a clear image on the screen. The difference between the refractive power of the jet and that of the air will cause an alteration in the distribution of light on the screen, which will make the projected areas lighter or darker than the surrounding space. Ether vapor poured from a beaker, or ether vapor vortex rings, may thus be made visible. If a Bunsen burner be placed in the cone of light a beautiful representation of the flame and the currents of heated gases will be formed on the screen.

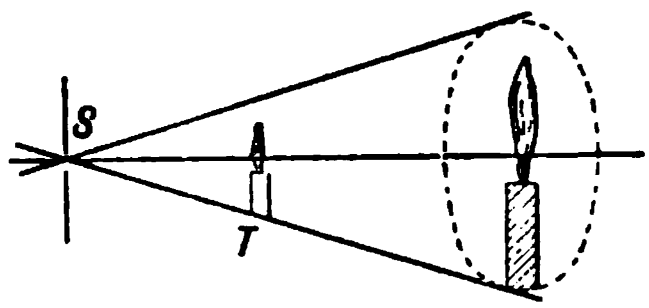


Fig. 540.

692. The Rainbow is a bright arc showing the spectral colors, due to the sunlight refracted by raindrops. Sometimes several bows are seen, the inner or primary bow being always the brightest, and all being arcs of circles with centers on the prolongation of the line passing from the sun through the observer.

The primary bow is violet on the inside, red on the outside; in the secondary bow the order of colors is reversed.

If parallel rays are incident on the upper half of a refracting sphere they will be in part refracted, internally reflected, and transmitted downward as shown in Fig. 541. Rays will also enter the lower half, and there will be multiple reflection within the sphere, but for the present we shall fix our attention upon the rays which reach the eye at O after one internal reflection. As indicated by the course of the rays incident at P, P_1, and P_2, there is an angle of minimum deviation, below which no rays once internally reflected pass. All the rays emerging at nearly this angle are parallel or nearly so, and therefore their intensity varies little with the distance from the drop, while rays emerging in other directions are widely divergent.

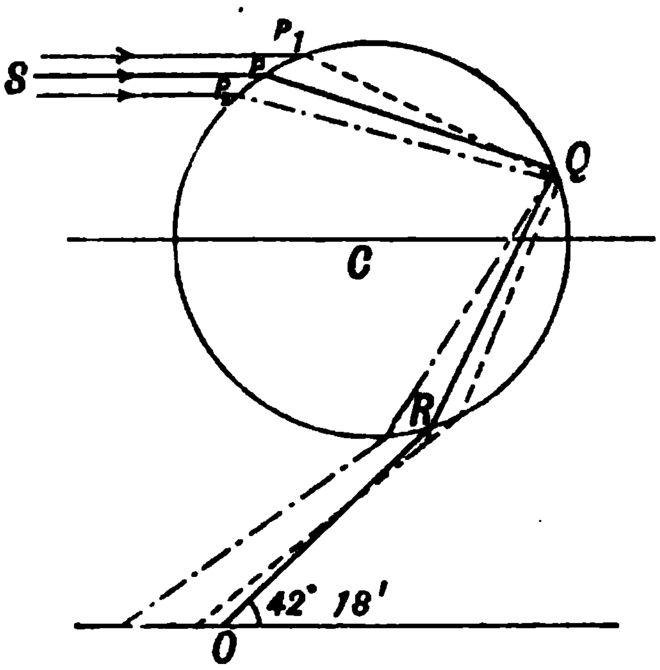


FIG. 541.

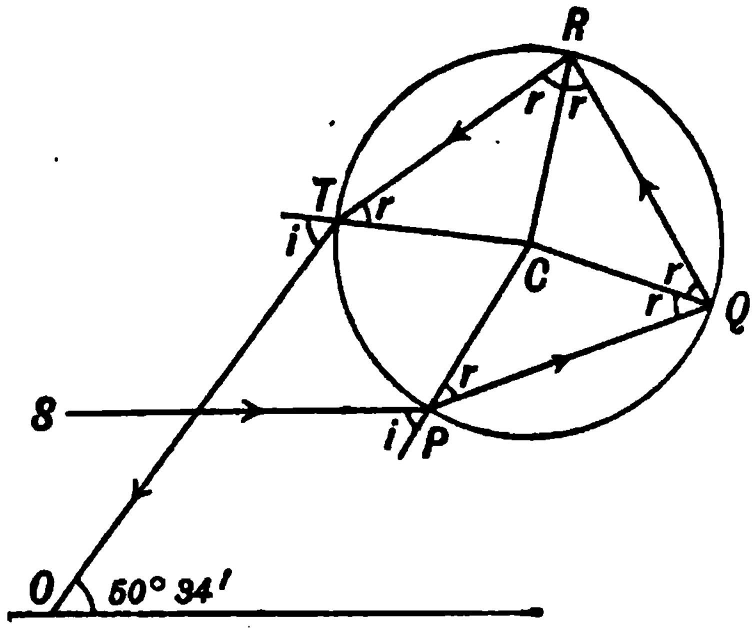


FIG. 542.

As n varies with the color, the minimum deviations are different for the various colors. In the primary bow the minimum deviation of the red is 137° 42′; of the violet, 139° 37′. In the secondary bow the corresponding angles are 230° 34′ and 233° 56′.

From Fig. 541 it appears that light will be received by the observer at O from all the raindrops lying in an arc subtending an angle 180°—D with the axis passing from the sun through the observer's eye. In the primary bow this angle is 42° 18′ for the red and 40° 23′ for the violet, so that the bow will be bordered with violet on the lower side, red on the upper. The secondary bow is due to rays incident on the lower half of the drop, twice internally reflected, and then transmitted downward, thus inverting the

order of the colors (Fig. 542). The angle subtended by this bow is D—180°, or 50° 34′ for the red, 53° 56′ for the violet.

An artificial rainbow may be made by causing a beam of sunlight to fall on a spherical vessel filled with water, through an opening in a screen. The interior of the circle reflected on the screen is illuminated by the scattered light which has been once reflected, while the space between the primary and secondary bow is quite dark.

INTERFERENCE.

693. Examples of Interference.—Fresnel, a young French artillery officer, about 1815, produced effects similar to those described in §648 by light diverging from a slit S and reflected from two adjacent mirrors MM inclined at such a great angle so as to be almost in the same plane. As shown by Fig. 543, the light arriving at any point P where the pencils overlap appears to come from the two virtual sources S_1 and S_2, the effect of which is precisely the same as that of the two real sources in Young's experiment. (The term virtual source applies to a point from which the waves appear to diverge without really originating at that point).

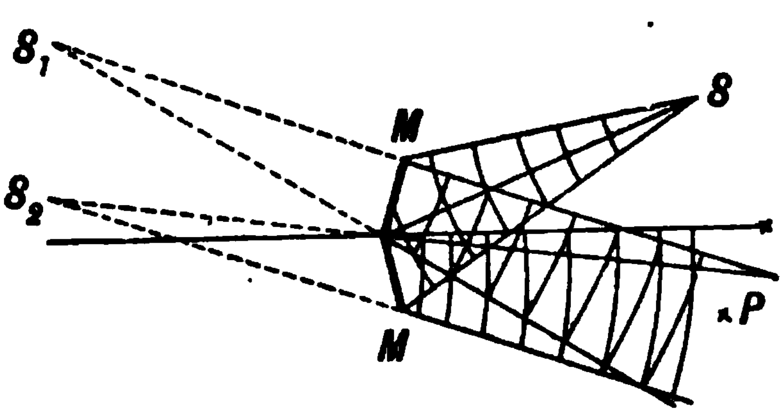


Fig. 543.

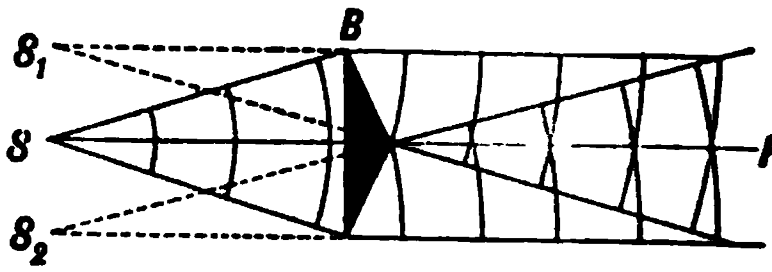


Fig. 544.

Another method of obtaining similar interference effects is by means of a convex lens L cut along a diameter in two halves which are slightly separated, giving two real or virtual images of the source, from which the waves diverge and overlap. This is known as the Billet split lens (Fig. 545).

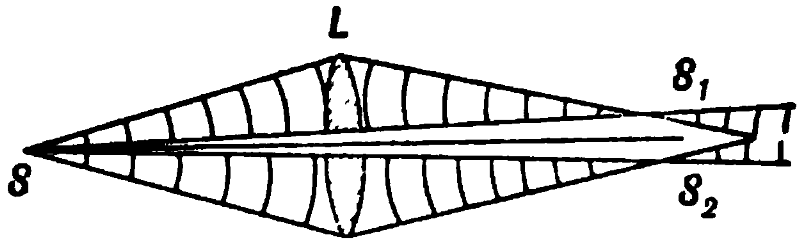


Fig. 545.

The interference effects due to Lloyd's single mirror (Fig. 546), are caused by waves coming respectively from the real source S and the virtual source S_1. The fringes are easily obtained by reflecting the light from a narrow slit or lamp filament at grazing incidence from a mirror of

black glass, in order that the effects may not be complicated by reflection from the rear surface.

Fresnel also produced interference effects by the use of a biprism B equivalent to two prisms of very small refracting angle placed base to base (Fig. 544). Here, again, it is evident that the transmitted light appears to come from the two virtual sources S_1 and S_2.

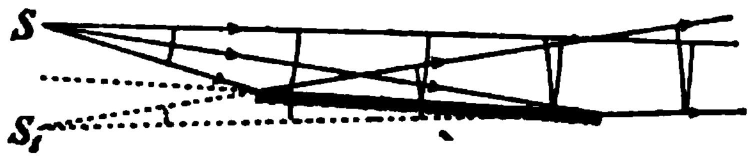


FIG. 546.

694. Newton's Rings.—Robert Boyle described the brilliant colors observed in soap bubbles and other thin films, an effect which appeared to depend solely on the thickness of the films, not on their nature. Newton investigated this phenomenon, which he tried, with poor success, to explain in terms of the emission theory. In order to secure a thin film of air, varying in thickness in a determinate manner from point to point, he pressed a convex glass lens of great radius of curvature against a piece of plane glass. If light falls normally on such a combination, light of a given color is found to be reflected in a greater proportion than the other colors from all points where the film has a given thickness, the predominant color varying with the thickness. As the loci of points of equal thickness form circles about the region of contact, colored rings are observed concentric with this point. These have been called Newton's rings, or the *colors of thin plates*. Colored rings are likewise observed in the transmitted light. These are not so brilliant, however, as those due to the reflected light, as the transmitted colors are mixed with a large proportion of unmodified white light. The colors in the two sets of rings are complementary—that is to say, the light transmitted through a given point is white deprived of the color which is most strongly reflected from that point. If monochromatic light is used the rings are alternately dark and of the color used. In a wedge-shaped film these bands are parallel to the edge of the wedge; in a film of uniform thickness circular bands are produced under certain conditions, uniform color effects under others. These colors of thin plates are seen in all kinds of thin transparent films, such as soap bubbles, films of oil on water, and thin sheets of mica.

695. Explanation of Newton's Rings.—Thomas Young showed that the colors of thin films can be very simply explained as a result of the interference of waves reflected from the two surfaces of the film, as shown below.

It has been found impossible to produce interference effects between two pencils from separate sources, or from different points of the same source. There is no permanent concordance in phase relations, amplitude, or direction of vibration. We need, therefore, consider only one point of the source at a time. The effects of adjacent points will be simply superimposed on those of the first, without mutual interference.

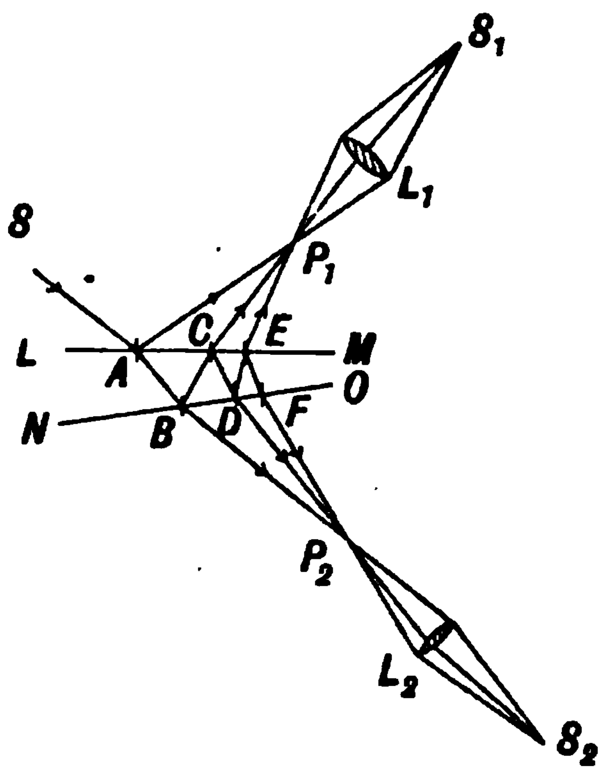


FIG. 547.

Let LM and NO (Fig. 547) represent two opposite elements of surface of the air film producing Newton's rings, slightly inclined to each other and so small that they may be considered plane. A narrow pencil from the point S of an extended source is incident at A, where a small part is reflected, the remainder being transmitted to B, where it is again subject to reflection and refraction. This process is repeated at C, D, E, etc., but the components become very weak after a few reflections. Owing to the inclination between the surfaces, the reflected pencils are not parallel, but intersect in the neighborhood of P_1, while the transmitted components appear to diverge from P_2. If the film increased in thickness toward the right, these points of intersection would be respectively on the opposite side of the film. The reflected and transmitted pencils may respectively be brought together again by the lenses L_1 and L_2, at the points S_1 and S_2. A maximum or a minimum of intensity may be produced at these points, according to the phase differences between the component rays. The eye or observing telescope must be focused on P_1 or P_2 to get the most distinct effect. If the film is very thin these points practically lie at its surface. If either the thickness of the film or the angle between its surfaces is large the pencils are so distant from each other or so divergent that they cannot all

enter the pupil of the eye. Under such conditions the bands become indistinct or vanish.

696. General Expression for the Difference of Path.—Let n be the index of refraction of the film, n_1, that of the surrounding medium, and for simplicity imagine the two surfaces to be parallel. To reach the wave front CP (Fig. 548) the light reflected from A travels the distance AP in the first medium while the interfering component has to travel the distance $AB+BC$ in the film. The wave would travel a distance n_1AP in air while traveling the distance AP in the medium of index of refraction n_1 and the distance $n(AB+BC)$ in the air while traveling the distance $AB+BC$ in the film (§678). Hence the equivalent difference of path in air, or the optical difference of path is

$$d=n(AB+BC)-n_1AP.$$

But $AC\sin i=AP$; $AC\sin r=CQ$; therefore

$$AP=CQ\frac{\sin i}{\sin r}=\frac{n}{n_1}CQ$$

hence

$$d=n(AB+BC-CQ)=n(AB+BQ)$$
$$=nAR\cos r=2nt\cos r.$$

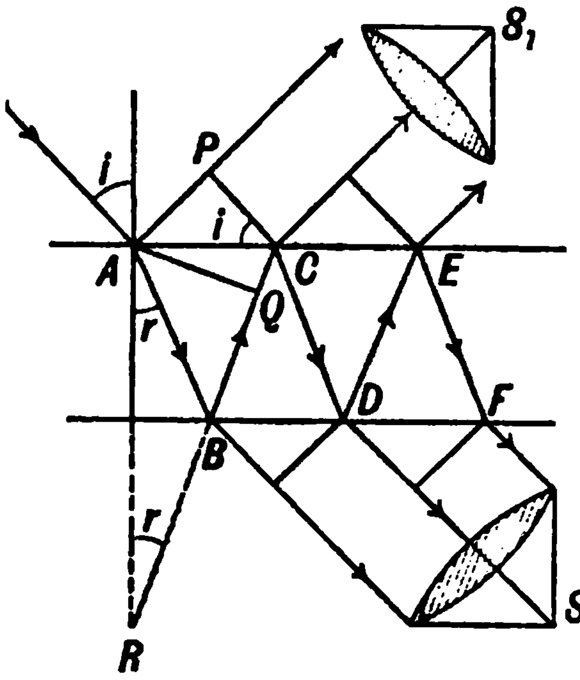


Fig. 548.

If the film is of air, $n=1$, $d=2t\cos i$ (if i is the angle of refraction in air). The effects at S_1 and S_2 are due to the superposition of all the components arising from multiple reflections within the film. Between successive pairs there is the same phase difference.

697. Phase Changes in Reflection.—(See §252.) Evidently, so far as geometrical differences of path are concerned, there should be reenforcement from the components reflected from the region of contact, where the thickness of the film is so small compared with the wave-length of light that it may be ignored. As a matter of fact, the center of the reflected system of fringes is black. Young inferred by analogy that at the boundaries of different media light waves are subject to changes of phase similar to those observed in the case of material waves §(252) so that waves incident from air on a more refracting medium may behave like waves of sound reflected from a medium denser than air, while a light wave traveling in the opposite direction will behave like sound waves emerging from the free end of an organ pipe. The waves reflected from the upper surface of the air film pass from a more to a less refractive medium; at the lower surface the contrary is the case. If t is small compared with the wave-length, there should be a difference of half a period introduced in the act of reflection, which will cause destructive interference. The transmitted components have a difference of phase of an entire period caused by two internal reflections, and therefore will be concordant. This would explain the black spot seen in the center of the reflected system of Newton's rings. It is also observed

that soap films as they get thinner run through a brilliant series of colors when viewed by reflected light, finally becoming black just before they break.

If the lens is of crown glass, the plate of flint glass, and if the interspace is filled with a liquid of intermediate index of refraction, such as oil of cloves, the central spot of the reflected system will be bright, that of the transmitted system dark. This confirms Young's theory.

When Newton's rings are produced by an air film, the condition for a maximum of given wave-length λ in the reflected light is (remembering that a loss of half a period in reflection is equivalent to a path difference of $\lambda/2$),

$$n\lambda+\tfrac{1}{2}\lambda=\frac{2n+1}{2}\lambda=2t\cos i$$

and for a minimum,

$$n\lambda=2t\cos i$$

In the transmitted light the maxima are given by

$$n\lambda=2t\cos i$$

and the minima by

$$\frac{2n+1}{2}\lambda=2t\cos i$$

In the above expressions n is the ordinal number of the rings counted from the center.

698. Film with Parallel Sides.—If the surfaces of a thin plate are perfectly plane and parallel the interfering rays are parallel, as shown in Fig. 548, and the eye or observing telescope must be focused for infinity to see the bands clearly. Since the difference of path between the components, $2nt\cos i$, varies with the angle of incidence, the phase relations will be different for rays reflected from different parts of the film, but will be the same for all rays reflected from the film at the same angle. The light from every point S_1, S_2, etc., of an extended source will be brought to separate points, S_1', S_2', etc., on the retina, so that there will be no overlapping of effects. The bands will in general be curved, their loci being given by $\cos i$ = constant. Such bands are sometimes known as Haidinger's fringes, or fringes of equal inclination.

699. Interference by Thick Plates.—It is usually impossible to get nterference effects by the use of a single wedge-shaped plate unless the inclination of the surfaces is very slight, because the interfering pencils will otherwise be too divergent to enter the eye simultaneously. If the surfaces of the plate are perfectly plane and parallel it is easy to obtain interference effects with monochromatic light with great differences of path between the components. The limit to the possible differences of path which may exist seems to be due to the lack of perfect homogeneity in the light from available sources, or to the probable fact that radiating centers emit detached wave groups corresponding to successive stimuli, these groups having different relations of phase, amplitude, and the direction of vibration, so that waves of one group cannot interfere with those of another. Consequently

the maximum difference of path which can exist cannot exceed the length of such a train of waves.

700. Stationary Light Waves.—Stationary waves (§253) may be expected if plane waves of light are reflected normally from a mirror, but as the distance between the nodal planes is only $\lambda/2$ and light waves are very short, it is difficult to verify their existence. Wiener did so by a very ingenious device. A glass plate AB (Fig. 549) was covered with a very thin photographically sensitive collodion film, and placed film downward over a silvered mirror MN with a very slight inclination between the two surfaces. After exposing the plate to a beam of light incident normally on the mirror it was developed, and dark bands were found in the film, running parallel to the line of intersection of the two surfaces, as indicated by the shading below MN. From the figure it is clear that the sensitive surface crossed the nodal and anti-nodal planes in such a manner as to produce this effect. At the points in contact with the mirror no effect was produced in the film. This proved the existence of a nodal plane at that surface, as in the case of the analogous sound experiment. This is the basis of a system of color photography invented by Lippmann.

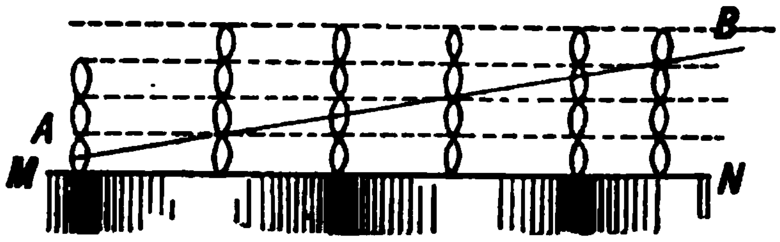


FIG. 549.

DIFFRACTION

701. If light from a small source or aperture passes by the edge of an obstacle and falls on a screen it is found that the illumination gradually fades away in the geometrical shadow, while outside the shadow a series of colored bands appears. If a card or knife blade is held between the eye and a distant source of light it will be found that the red light is most deflected into the shadow, the violet the least, so that a short spectrum is formed. Such phenomena are examples of what is known as **Diffraction.** They are, in fact, interference phenomena between wavelets coming from adjacent points of the same wave front.

Let us find the effect of an extended plane wave front AB (Fig. 550) at the point P. In accordance with Huyghens' principle, the resultant effect at P may be regarded as the sum of the effects separately due to all the points in the wave front, each originating its independent set of wavelets. Waves of different lengths must be separately considered in this analysis. If $OP=r$, describe about P as a center spheres of radii $r+\lambda/2$, $r+\lambda$, $r+3\lambda/2$, etc.

These spheres will intersect the wave front in circles, as shown in Fig. 551, concentric with O, the *pole* of the wave with respect to P. The areas between successive circles are called ***half-period zones.***

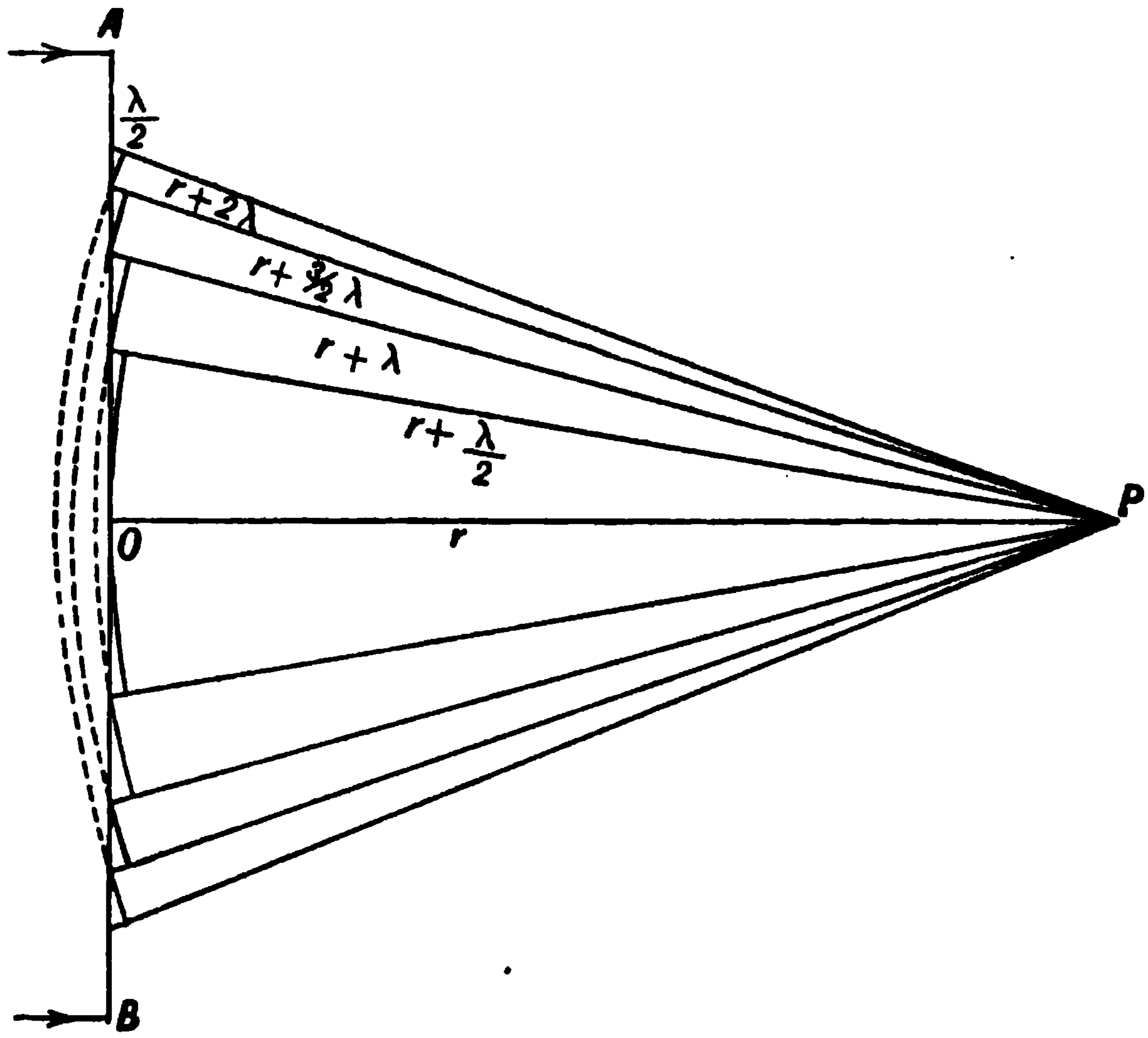


Fig. 550.

The student may easily find by calculation that these areas are approximately equal. It is evident that the disturbances originating in all points in a circle about O will reach P at the same time, and that the average phases of the resultant effects at P of successive zones will differ by half a period. Although the areas of the zones are practically the same, the amplitude produced by each at P slowly diminishes as its radius increases, on account of increasing obliquity and distance with respect to P. The total amplitude produced by the wave at P is, therefore, the algebraic sum of a series of terms slowly diminishing in magnitude and alternating in sign (direction of displacement). If a_1, a_2, a_3, etc., are the amplitudes at P due to the central area and successive zones, and A the resultant amplitude.

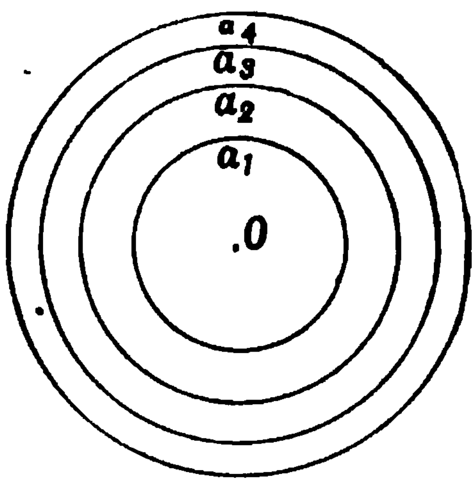


Fig. 551.

$$A = a_1 - a_2 + a_3 - a_4 + a_5 \cdots \pm a_n$$
$$= \tfrac{1}{2}a_1 + \tfrac{1}{2}(a_1 - 2a_2 + a_3) + \tfrac{1}{2}(a_3 - 2a_4 + a_5) \cdots \pm \tfrac{1}{2}a_n$$

As the successive terms differ very slightly from each other and diminish in accordance with a regular law, the quantities in parentheses and a_n may each be placed equal to zero. Therefore,

$$A = \tfrac{1}{2}a_1$$

or the amplitude at P due to the whole wave is one-half and the intensity one-fourth that due to the central element if it alone were effective. If the whole wave except the central element

FIG. 552.

is covered the illumination at P will be actually increased, the amplitude in that case being a_1. If all but the two central elements are covered the effect at P is $A = a_1 - a_2 = 0$ nearly. If three elements are uncovered, $A = a_1 - a_2 + a_3 = a_1$ nearly. These conclusions are easily verified by experiment. If small circular openings of different sizes be placed in a pencil of light diverging from a pinhole, maxima and minima will be found in the centers

of the bright areas projected on a screen through the openings (Fig. 552). These holes decrease regularly in size from 1 to 9. If the screen be moved (thus changing the number of effective half-period elements subtended by the holes at the screen) maxima change to minima and *vice versa*, or if white light is used the bright spot at the center changes color. The central spot is surrounded by a series of colored bands of similar origin, but not so easily explained by elementary methods. If a hole is smaller than the first two half-period elements, there are no maxima and minima within the illuminated area on the screen, as there can be no possible disordance of phase in the wavelets coming through the hole, and consequently a diffuse circular patch of light is cast on the screen, which increases in size as the opening is made smaller.

FIG. 553.

If a small disk be placed in the path of the light, so as to cover a few half-period elements as viewed from P—say three—the amplitude at P will be $A = a_4 - a_5 + a_6 - a_7 \cdots = \frac{1}{2}a_4$. A bright spot will therefore be seen at the center of the shadow, nearly as intense as though the disk were removed. At adjacent points not on the axis there will be discordance of phase between the disturbances coming around the edge of the disk, resulting in destructive interference.

It is easy to perform this experiment by mounting a perfectly circular disk several millimeters in diameter on a piece of glass plate and placing it in the pencil of sunlight from a small pinhole opening several meters away. The bright spot in the center of the shadow may then be seen on a screen a few meters beyond the disk. A reproduction of a photograph of this effect is shown in Fig. 553.

FIG. 554.

702. Waves through Large Opening.—The points a, b, c, etc., in the wave front AB (Fig. 554) act as centers of disturbance and propagate wavelets to the tangent surface $A'B'$. It is evident from the figure that only those wavelets between the lines AA' and BB' conspire to form a common wave front. Outside of these lines the wavelets cross each other in all directions and in all possible phases at random, so that the resultant disturbance is zero except in the immediate neighborhood of AA' and BB' where diffraction effects are produced.

703. Waves through Small Opening.—From Fig. 555 it is clear that no opposition of phase between the elementary wavelets from a, b, c, etc., can arise until the point P_1 is reached, where the difference between AP_1 and BP_1 is half a wave-length, and even then the disturbances from the extreme points A and B alone are in opposite phase. Only when this difference of path is a whole wave-length can complete destruction arise. In this case we see that the disturbances from two halves of the opening reach P_1 with an average difference of path of half a wave-length, so that the wavelets nullify each other pair by pair. If the opening is less than a wave-length in width some effect is produced even at the point P_2. The effect is evidently always greatest at P_0, where the wavelets meet very nearly in the same phase, and least at P_2 where there is the greatest diversity of phases.

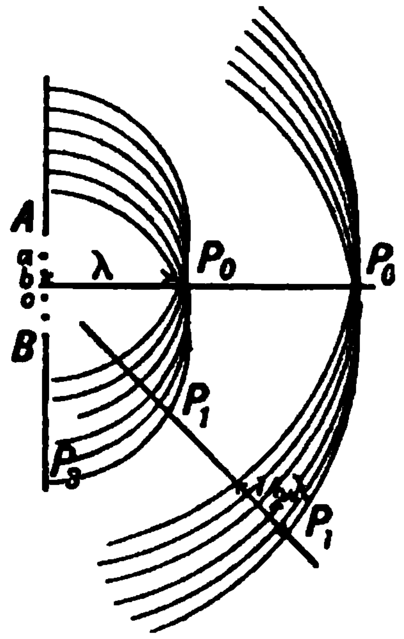


FIG. 555.

704. Narrow Slit.—If two straight edges are opposed so as to form a narrow slit of width AB (Fig. 546) there will be a bright band at P_0 if only the two central half-period elements of an incident wave are exposed. If two on each side are exposed the effect at P_0 is

$$A=2a_1-2a_2 \text{ (nearly zero)}$$

If three half-period elements on each side are exposed

$$A=2a_1-2a_2+2a_3 \text{(maximum)}$$

Thus there will be successive maxima and minima at P_0 as the slit is widened. If the slit subtends two or any even number of half-period elements as viewed from P_1, apoint off the axis (T being its pole), they will neutralize each other in pairs; if it subtends an odd number of such elements, there will be destructive interference between pairs, leaving the odd one effective, consequently there will be a series of maxima and minima on each side of the axis.

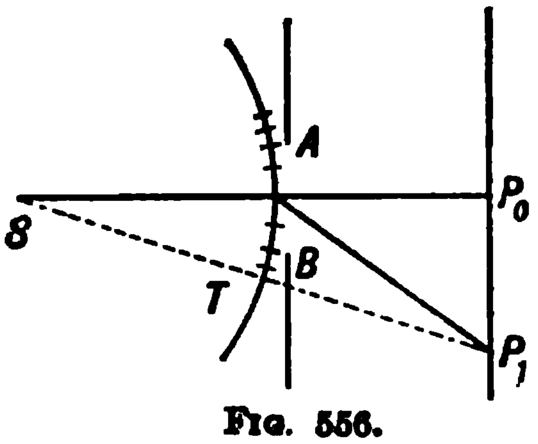


FIG. 556.

705. Width of the Bands Formed by a Narrow Slit.—If $AP_1-BP_1=\lambda$ (Fig. 557) we may consider the effects at P_1 of AO and OB to be nearly the same numerically, but to differ in average phase by half a wave-length. The two cancel each other. At P_2, where $AP_2-BP_2=\frac{3}{2}\lambda$, we may imagine the slit divided into three nearly equal strips, which contribute effects at P

alternating in phase. Two cancel each other, leaving the third effective. If D is the distance of the screen from the slit, the width of the central maximum is $2P_0P_1$, and it can easily be shown as in §648 that

$$2P_0P_1 = \frac{2D\lambda}{AB}$$

The other bands are of half this width, or $D.\lambda/AB$. The width of all the bands is, therefore, inversely proportional to the width of the slit. The

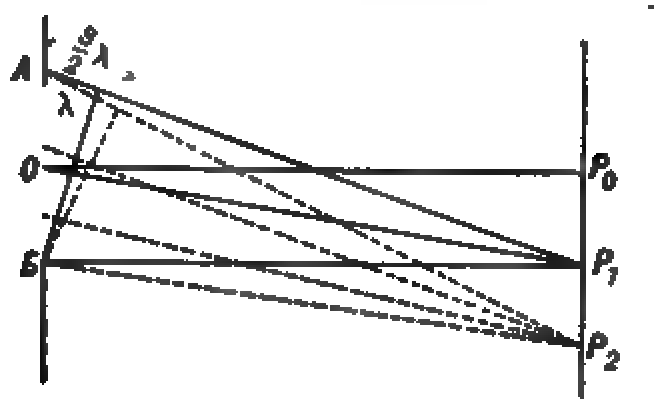


FIG. 557.

central maximum is white, the other are narrow spectra bordered by violet on the inside, red on the outside (since P_0P_2 for violet is less than the corresponding distance for red.) These effects may be observed by allowing light from a narrow slit to pass through a second adjustable slit and fall on a screen, or more simply by looking through a narrow slit or the space between two fingers at a distant light.

FIG. 558

Within and close outside the shadow of a wire or needle cast by a linear source similar fringes are observed. (See Fig. 558, showing shadows of needles of different sizes.)

706. Resolving Power.—If the light from a narrow slit passes through another slit to a screen the central maximum may be regarded as an image

of the first slit (corresponding to a pin-hole image). The wider the second slit is opened (up to the point where diffraction effects cease) the narrower and sharper this image will be. Similar considerations apply to light from point sources through circular openings. If we look through a small pinhole at a distant light it will appear much larger than when viewed with the naked eye. The filament of a lamp appears thicker when seen through a narrow slit. If an image is formed by a lens or mirror the same conditions hold as for a narrow slit, the lens or mirror preserving the uniformity of phase of the whole wave with respect to the focus that exists for a narrow slit with respect to its central maximum. Consider the image of a narrow source S (Fig. 559). At P_1 and P_2 on each side of P_0 there will be a minimum

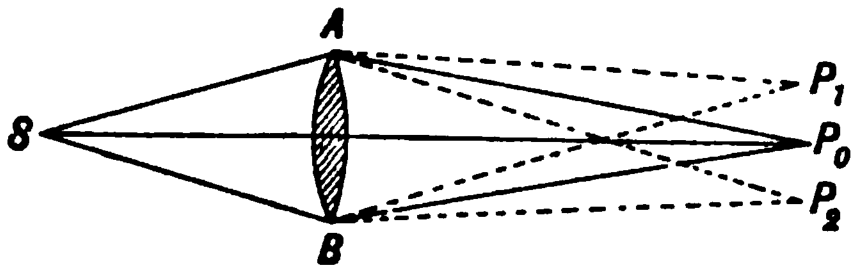


FIG. 559.

if $AP_2-BP_2=\lambda=BP_1-AP_1$ in which case the disturbances from the two halves of the lens reach P_1 and P_2 in opposite phases and cancel each other. The width of the image, which is merely a diffraction maximum, is, therefore, P_1P_2. If the source is a point, the central maximum will be of the shape of the opening, but differently oriented, because any particular dimension in the source will be inversely proportional to the same dimension in the opening, as appears from the relation (§705).

$$P_1P_2=2P_0P_1=2D\lambda/AB$$

Observation shows that two diffraction maxima cannot be clearly separated if they are closer than the distance from a maximum to the adjacent minimum. If the image of S for example lies at P_1 it can barely be seen as separate from P_0. The two images will overlap if the angle subtended by the objects at the lens is less than $\alpha=P_0P_1/D=\lambda/AB$. This is called *the angle of minimum resolution.*

Such conditions apply only to small sources or objects. If the source is large each point will have a diffraction maximum at the focus. These maxima will overlap and blot out diffraction effects except at the boundaries of the image.

Stars are practically point sources of light. Their images when formed by a telescope with a small objective appear much larger than when formed with a large telescope, the diameters of the central maxima being inversely proportional to the diameter AB of the lens. The image of a double star formed by a small telescope may be one large blur, while that formed by a large telescope consists of two distinct points of light. The ability to separate the images of two small adjacent sources is called resolving power, and may be shown to be directly proportional to the diameter of the lens, mirror, or prism forming the image—or to the cross-section of the effective beam of light, if it does not cover the above.

If we had larger eyes we could see much finer details than we now do. On the other hand, if we look through a small pin-hole at a distant light it will appear much larger than when viewed with the unaided eye. For a similar reason it is physically impossible for small insects to see details clearly. To them an incandescent lamp filament must appear as it does to us when we look at it through a very small pin-hole.

707. Diffraction Grating.—If there are a number of narrow and equidistant parallel openings in a screen, each pair of openings will produce effects similar to those observed in Young's double slit experiment. If a lens is placed in front of such a diffraction grating, as it is called (Fig. 560), the same path difference will

FIG. 560.

exist between any pair of adjacent parallel rays. If $a = AB$ is the distance between openings and if the angle between OP_0 and OP_1 is θ, the difference of path between corresponding parts of adjacent slits (*e.g.* A and B) is $AC = a \sin\theta$, and the condition that there shall be a maximum at P_1 for the wave length λ is

$$a \sin\theta = n\lambda$$

If light of one wave-length is used, there is a series of maxima on either side of the axis in positions where $a \sin\theta$ equals 1, 2, 3, etc., wave-lengths.

If white light is used, corresponding maxima for two different colors are at different distances from the axis. The central maximum P_0 is white, as the condition for reënforcement at that point ($n = 0$) is the same for all colors. The other maxima are drawn out into spectra on each side of the axis, as θ varies

with the wave-length. The value of the ordinal number n determines the *order* of the spectrum.

If θ is a small angle the distances between points in the spectra are nearly proportional to the differences of the corresponding wave-lengths, so that the spectra formed by gratings are said to be *normal*, as contrasted to those due to prisms, in which there is no simple law of distribution. All gratings give spectra which are alike in their distribution of colors, although they may differ in length. The lengths of the spectra increase directly as the order of the spectrum, so that those beyond the first overlap, and they also rapidly diminish in intensity.

A grating such as has been considered above is called a *transmission* grating and consists of lines ruled (several thousand to the inch) on glass by a diamond point. *Reflection* gratings are made by ruling lines on a polished surface of speculum metal; the incident light is reflected by the polished strips between the rulings. Such gratings are ruled by entirely automatic machines.

The effect due to a grating is precisely the same, so far as position of maxima is concerned, as that due to two slits with the same interval between them. The intensities of the grating spectra, however, are far greater, the amplitude of vibration being in proportion to the number of openings.

The resolving power of a grating is also greater. The width of a maximum in the interference bands given by two slits is (§648) $w = D\lambda/a$. The width of the maxima given by a grating having N openings is $w = D\lambda/Na$, since N is the aperture of the grating, so that this width is inversely as the breadth of the grating (§648). Diffraction gratings are generally used for measurements of wave-length.

A grating with crossed lines gives a beautiful series of crossed spectra. This effect may be observed by looking through a handkerchief or umbrella top at a distant light. Brilliant diffraction effects are also obtained by looking at a source through a cobweb or feather, or from the light reflected from mother of pearl. In the latter case the effect is due to striations, as may be proved by transferring the effects to wax by pressure.

OPTICAL INSTRUMENTS AND MEASUREMENTS

708. The Eye is an essential part of any optical combination. Like a photographic camera, it is a closed chamber into which light can enter only through the lens. As the camera lens throws

an image on a sensitive photographic plate which excites the silver grains, the lens of the eye forms a picture on the mat of sensitive nerve endings covering the retina. The amount of light entering the camera is regulated by an "iris" diaphragm of adjustable size; similarly the amount of light entering the eye is controlled by the size of the pupil, which automatically changes in diameter between the limits of about 2 and 5 mm. The parts of the eye are shown in Fig. 561. *S* is the sclerotic membrane, the outer enclosure of the eye. *C* is the cornea, a strong transparent membrane. *I* is the iris, the colored part of the eye, with a central orifice, the pupil, which admits light through the crystalline lens *L*, which focuses images on the retina *R*. The nerve endings covering the retina run together like the strands of a cable into the optic nerve *O*, which conveys stimuli to the brain. Muscles attached to the periphery of the lens can by their contraction or relaxation so change its curvature as to enable it to focus either distant or very near objects on the retina. This process is called accommodation. Two objects are clearly seen separately when the angle between them at the eye is a little less than 1,′ or the distance between the retinal images is 0.005 mm. Details are, therefore, more clearly seen as an object is brought nearer, as the angle subtended by it and the size of the retinal image are then larger; but there is a limit to the power of accommodation of the eye, so that usually no object nearer than about 25 cm. can be clearly seen. This is called the distance of most distinct vision. The normal adjustment of the eye when at rest, is for "infinity," as may be verified by suddenly raising the eyes when they have been unemployed and looking toward distant objects. They will be in focus.

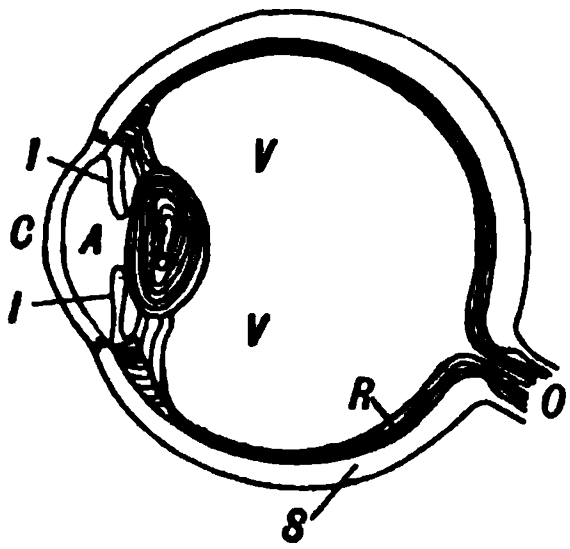


FIG. 561.

Between the cornea and the crystalline lens is the aqueous humor, *A*, and between the lens and the retina is the vitreous humor *V*, both transparent fluids with a mean index of refraction equal to 1.336. The lens is built up of transparent horny layers, increasing in density, hardness, and refractive power toward the

center. The index of refraction of the outer layer is 1.405; of the next, 1.429, and of the central region 1.454. The average index of refraction is about 1.437. This increase in density toward the axis serves partly to correct spherical aberration, which is also diminished by the iris diaphragm.

Objects such as printed letters can be clearly seen through a pin-hole in a card, even if they are as close as 2 cm. to the eye. This has been attributed to an over-correction of the lens for spherical aberration, so that a narrow pencil passing through the axis of the lens has a very short focus. It is obvious that so much over-correction would be worse than no correction at all. As a matter of fact, a pin-hole image is formed on the retina, the lens merely sharpening the effect. The fact that the apparent size of the object varies as the card is moved back and forth, the object remaining at rest, shows that the image is due mainly to the pin-hole.

709. Vision.—The retina is covered, except over the optic nerve, by a large number of very small fibrous bodies, "rods" and "cones," nerve endings which are in some way stimulated by light waves. Over the optic nerve is the "blind spot," so called because if the image falls on this part of the retina it ceases to be visible. By closing one eye and looking steadily with the other at one of two small objects about two inches apart, a distance may be found at which the other object will disappear. Excitation of the optic nerve lasts about one-tenth of a second after the stimulus ceases, so that if intermittent stimuli are applied at intervals less than this a steady effect is produced. This is called *persistence of vision*. The trail of the lighted end of a cigar if it be rapidly moved and the apparent continuity of moving pictures depend on this effect.

Sometimes the normal spheroidal shape of the lens is altered so that the curvatures are not the same in different planes. Light from a point will then pass through the eye as an astigmatic pencil with two focal lines instead of a point image (§685). Horizontal and vertical lines at the same distance cannot be simultaneously brought into focus. Such eyes are said to be astigmatic. Other defects arise from change of curvature or from loss of the power of accommodation. If eyes are short sighted, the principal focus falls short of the retina, and distant objects cannot be clearly seen. If they are long sighted, the principal focus is on or near the retina, and images of near objects cannot be formed

on the retina. For the first defect concave spectacles are the remedy; for the second they must be convex.

In normal eyes the nerve endings on which fall corresponding points of the two retinal images lead to the same nerve centers, so that the two pictures are exactly superimposed and a more intense effect secured than with one eye alone. If one eye-ball be forcibly twisted out of position double images will be seen. A further advantage given by two eyes is that an object is viewed from two slightly different directions, which gives the impression of relief. This principle is applied to the stereoscope, in which two photographs taken from slightly different points of view are viewed by each eye separately. The two images will be superimposed in such a manner that the object appears to stand out in space.

With two eyes it is also easier to estimate distances than with one. There is an angle between the two lines of sight to the object, which the brain unconsciously estimates. In general the sizes of objects are inferred from their angular magnitudes and estimates of their distance based on experience, or by comparison with adjacent objects, such as trees and houses, the sizes of which are approximately known. Such estimates are influenced by the clearness with which details are seen. In places where the atmosphere is unusually clear, as in Arizona, this leads to the underestimation of distance. Conversely, objects seen in a fog appear to be more distant than they are, owing to the indistinctness of their details. The angle subtended by them, however, corresponds to the actual distance, hence they loom larger than they are.

710. Irradiation.—This is the apparent increase in size of objects as they become brighter. The crescent of the new moon, for example, looks larger than the remainder of the disk, the "old moon," which is illuminated by the earth alone. The filament of an incandescent lamp appears to increase in size as it passes from ordinary temperatures through red and white heat. This effect was long supposed to be due to the spreading of the retinal image on account of stimulation of nerves outside of its boundaries, in much the same way that an overexposed photographic image is affected. It is now believed by some that the effect is due merely to spherical aberration of the eye, which becomes more noticeable as the intensity of the source increases.

711. The Simple Microscope or magnifying glass is a single convex lens through which objects at or within the principal focus of the lens are viewed. As shown in Fig. 562, an enlarged virtual image $A'B'$ is formed subtending at the lens the same angle α as the object AB. The linear size of this image is determined from the relation

$$I = (v/u)O$$

As the normal adjustment of the eye is for infinity, the object is usually at or very near the principal focus. In no case can the image be clearly seen when nearer than the limit of distinct vision. The actual linear magnitude of the image counts for little; the size of the retinal image depends on the angle subtended at the eye, and if the latter is very near the lens this angle is substantially that subtended from the lens. The lens simply increases the power of accommodation of the eye, so that the object may be brought nearer and thus subtend a greater

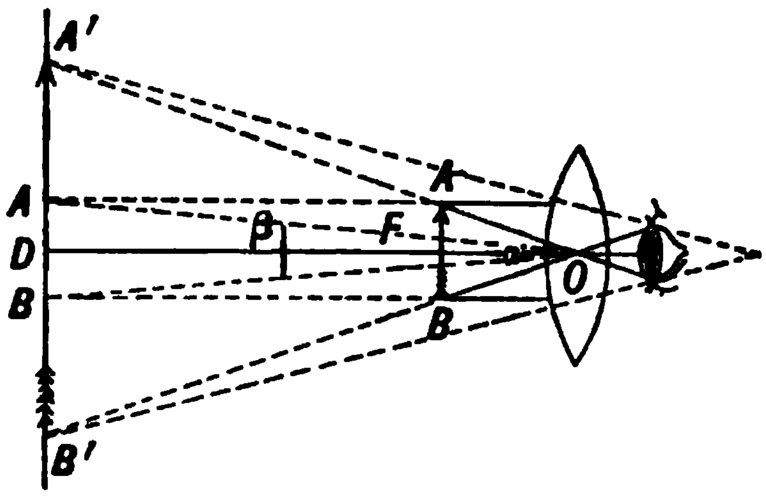


FIG. 562.

angle. With the unaided eye, the greatest detail is observed at the distance of most distinct vision (25 cm.) where it subtends the angle β (Fig. 552.) With the lens, the object is brought nearer, approximately to the principal focus, and the angle subtended by it increases from β to α. The magnification M of the retinal image is, therefore, α/β. If f is the focal length of the lens, d the limit of distinct vision,

$$AB = 2d.\tan \beta/2 = 2f \tan \alpha/2$$

Therefore, if these angles are small

$$M = \alpha/\beta = d/f$$

712. Power of Lenses.—The magnifying power of a lens is, as shown above, inversely proportional to the principal focal length, hence $1/f$ is a measure of its power. The practical unit of lens power is that of a lens with a focal length of one meter. This unit is called a *diopter* or *dioptric.* The power of converging lenses is considered positive, that of diverging lenses negative. The relation deduced in §686 shows that the power of a number of lenses in contact is the algebraic sum of their individual powers.

713. Eye-pieces.—The part next the eye of an optical train of lenses, such as those of telescopes and compound microscopes, usually consists of some form of simple microscope known as an eye-piece. With a single lens, much of the light from the real image formed by the objective O, which is usually viewed through the eye-piece, would be lost. In order to avoid this, light is gathered in toward the axis by a second lens, called the field lens F (Fig. 563). Nearly all the light would pass by the edge of the eye lens E if F were absent. It may be shown that a combination of two lenses of the same kind of glass is nearly achromatic if they are placed at a distance from each other $d=(f_1+f_2)/2$. This property is utilized in most eye-pieces which consist of a field lens and eye lens.

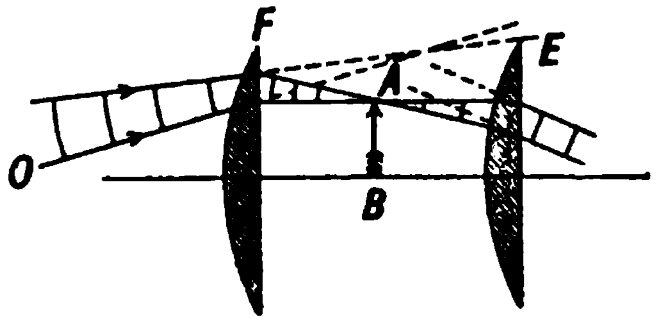


Fig. 563.

In Huyghens' eye-piece $f_1=3f_2$ (Fig. 563). Hence $d=2f_2$ and if the image due to the objective and the first lens is formed half way between the lenses the emergent light will be parallel and a virtual image formed at infinity. If a cross thread is used, it must be placed at AB. The lenses are convex toward the incident light and of such curvature as to reduce the spherical aberration to the minimum.

In the Ramsden eye-piece (Fig. 564) $f_1=f_2$. If the lenses are placed apart at the distance $(f_1+f_2)/2$ dust particles on the field lens would be visible through the second. In order to avoid this, the lenses are usually placed at a distance of $2f/3$. The principal focal point of the combination is at a distance $f/4$ in front of the first lens. The object, or the real image due to the objective, is at this point, and the final virtual image is at infinity. The chromatic aberration is small, and the spherical aberration is reduced by using plano-convex lenses with convex surfaces facing each other.

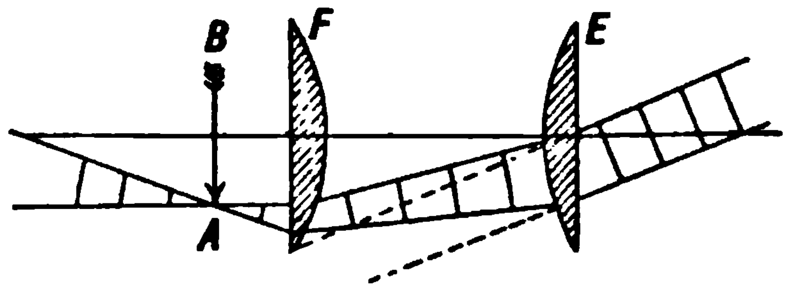


Fig. 564.

In all these eye-pieces the emergent red and violet rays are nearly parallel, hence the virtual images formed by the different colors subtend very nearly the same angle at the eye, and are, therefore, of the same size, but not quite equally sharply focused on the retina.

714. Compound Microscope.—In order to extend the limit of magnification beyond the point obtainable with a single microscope, a combination of lenses is used. An enlarged real image

$A'B'$ (Fig. 555) is formed by an object lens or train of lenses, and this image is further enlarged by an eye-piece, such as that of Huyghens, used as a simple microscope, which gives a virtual image $A''B''$. The front lens of the objective train is usually of the hemispherical form described in §683, which has a great angular aperture, with very little spherical aberration. There are in addition a number of other lenses of different shapes and kinds of glass, so combined as to reduce spherical and chromatic aberration to a minimum and to give a plane focal surface. A typical combination is shown in Fig. 566.

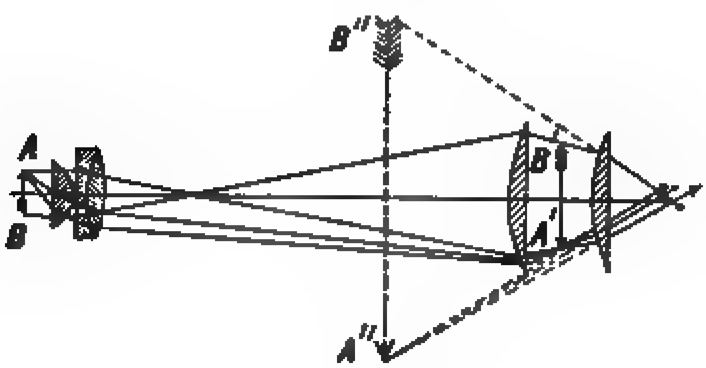


FIG. 555.

FIG. 566.

The magnifying power of the objective, of focal length f_1, is

$$M_1 = I_1/O = v_1/u_1$$

That of the eye-piece is, as shown in §711,

$$M_2 = I_2/I_1 = d/f_2$$

where d is the minimum distance of distinct vision. The magnification due to the combination is

$$M = M_1M_2 = I_2/O = Ld/f_1f_2 \text{ approximately}$$

where L is the distance between the objective and the eye-piece.

The minimum distance between two small objects A and B seen through a microscope which will permit of clear separation of their diffraction images is obtained by a slight modification of the expression found for the minimum angle of resolution, $\alpha = \lambda/AB$ (§706). The minimum value which d can have is thus found to be $\lambda/2$, when the object is at the surface of the lens. Since this distance is proportional to the wave-length, details which may be clearly seen when the object is illuminated by blue light will be indistinct when red light is used.

715. Astronomical Telescope.—The object glass of a telescope forms a real and of course greatly reduced image $A'B'$ of a distant object (Fig. 567). The object and its image subtend the same angle α at the objective, and the object subtends practically the same angle α if viewed directly by the eye. If, however, the eye views the image formed by the objective at the distance of most distinct vision, (from the point E), this image will subtend an angle β which is larger than α, and the apparent magnification is $M_1 = \beta/\alpha$. When this image is viewed through

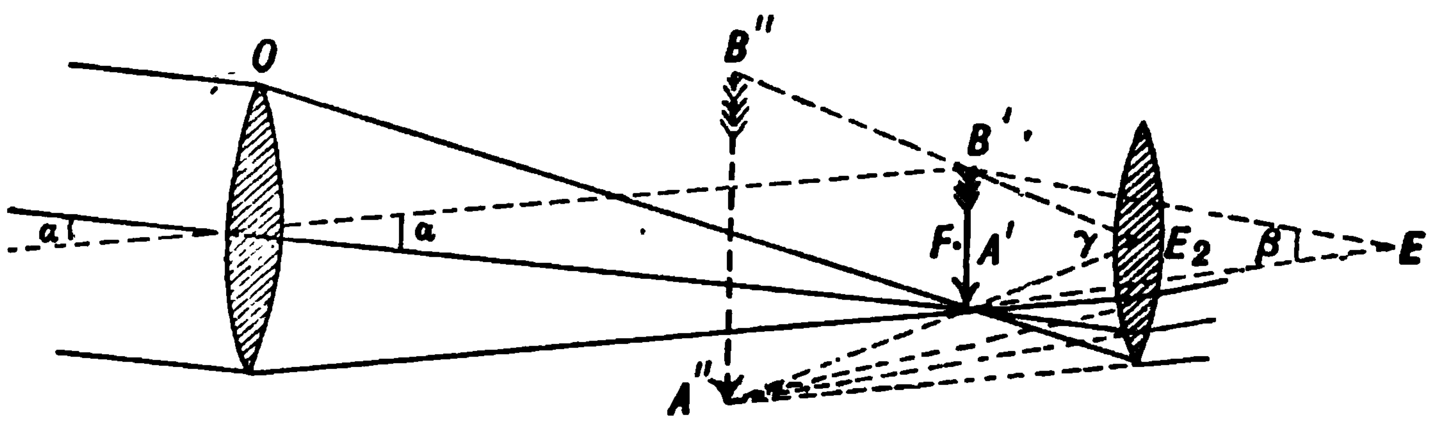


FIG. 567.

an eye-piece, (the eye now being at E_2), there is further magnification, the image subtending the larger angle γ. The magnification due to the combination is

$$M = M_1 M_2 = \frac{\beta}{\alpha} \times \frac{\gamma}{\beta} = \frac{\gamma}{\alpha} = \frac{f_1}{f_2}$$

The limiting angle of resolution between two linear sources is proportional to λ/A, where A is the diameter of the objective (§706).

For astronomical purposes there is no disadvantage arising from the fact that an inverted image is formed by a telescope, but when the instrument is to be used for terrestrial purposes it is necessary to add an additional lens or pair of lenses to reinvert the real image formed by the objective. This adds inconveniently to the length of the tube. If the image is inverted by reflection from a combination of prisms the length may be diminished, but for most purposes where only small magnification is required the form of telescope devised by Galileo is most convenient.

716. Dutch or Galilean Telescope.—This type is used for opera glasses and for marine glasses. As shown in Fig. 568 an

erect virtual image is formed, the magnification being $M = f_1/f_2$ (§715). The tube has a length approximately equal to the difference between the focal lengths of the objective and the eye-piece, while in the ordinary telescope the length is the sum of these distances.

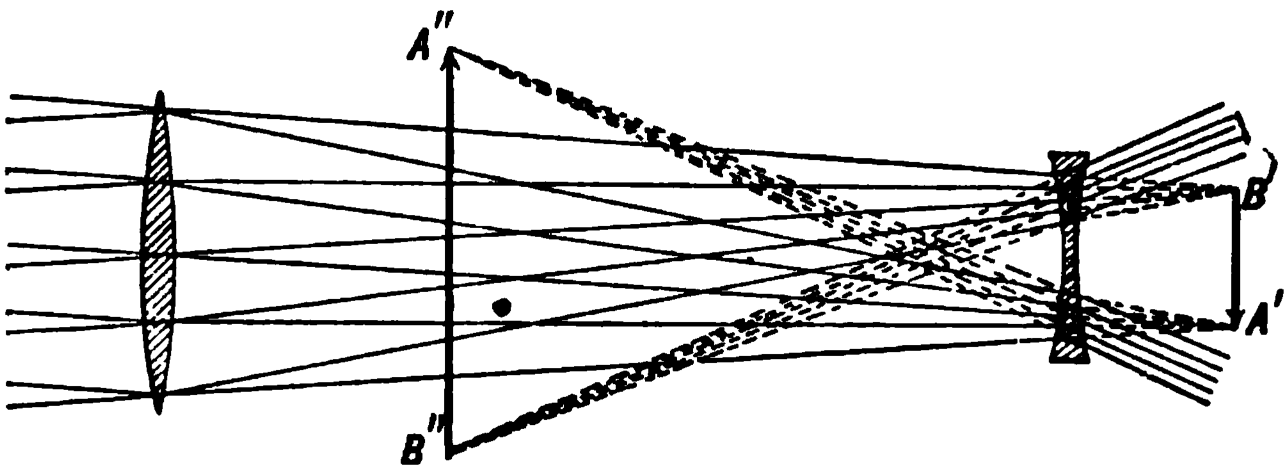


FIG. 568.

717. Reflecting Telescope.—The objective lens may be replaced by a large concave mirror. In this way chromatic aberration may be entirely avoided, but spherical aberration is more troublesome than with refractors. As the real image is formed along the axis of the mirror and in the path of the incident light, special devices are necessary in order to view it. In the Newtonian telescope the image is reflected to one side by a small right-angled prism which cuts off very little light, and is viewed by an eye-piece in the side of the tube. Herschel tipped the mirror slightly so that the image was formed at the edge of the open end of the tube, at which point the eye-piece was fixed. In other forms a small mirror in the axis reflects the image back into an eye-piece set in the center of the objective itself, so that it can be viewed from behind.

718. Photographic Camera.—This is a form of camera obscura in which the image formed by a lens falls on a sensitive photographic plate. The requirements demanded for the lens are exacting and in some cases contradictory to each other. It must give images free from spherical and chromatic aberration, and in many cases have great light power and a large field of view. The focal surface must be plane, and the magnification must be the same in all parts of this plane, so that no distortion is produced. The depth of focus must be great, that is, objects at different dis-

tances must have images approximately in focus at the same time on the plate. As the film is most sensitive for the shortest waves, the lenses must be corrected for the violet and the yellow, instead of blue and red. A diaphragm with a small opening is used in front of the lens, if it is a single achromatic combination, such as is used for landscape work. This reduces spherical aberration and at the same time gives a greater depth of focus (approximating to the principle of the pin-hole camera, in which the focus is nearly independent of the distance). A diaphragm

FIG. 569. FIG. 570.

with a single lens results in a distorted image, however, as shown in Fig. 569, which represents the distortion of the image of a quadrilateral network with the diaphragm in front of the lens, and Fig. 570, which gives the effect due to a diaphragm behind the lens. The cause is readily seen to be due to the difference in deviation of pencils passing through the center and the edge of the lens respectively. If two lenses are used, with the diaphragm at the optical center of the combination, these distortions correct each other.

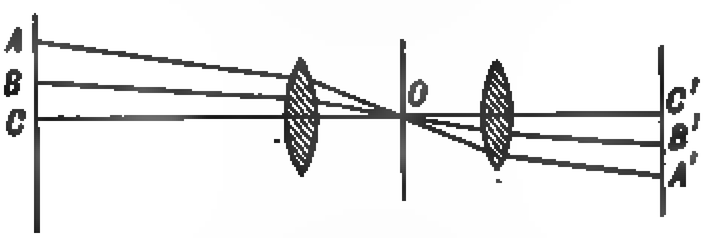


FIG. 571.

As shown in Fig. 571, the perfect symmetry of the incident and the transmitted secondary axes AA', BB', CC', etc., with respect to the opening O shows that the distances AB, BC, etc., in the object are in the same ratio as the corresponding distances $A'B'$, $B'C'$, etc., in the image, so that there is no distortion, and if A, B, C are in the same plane, $A'B'C'$, etc., must be in the same plane. Such lenses are called *rectilinear or orthoscopic doublets.*

The size of the photographic image of a distant object is nearly propor-

tional to the focal length of the lens. It is, however, inconvenient to give a great length to the camera box. This difficulty is avoided by the use of the *teleobjective*, in which a concave lens L_2 is placed behind the converging lens L (Fig. 572.) The divergent effect of this lens gives a virtual focal length equal to PF, while the camera box has the much smaller length LF. A greatly enlarged image is secured, but the field of view is reduced.

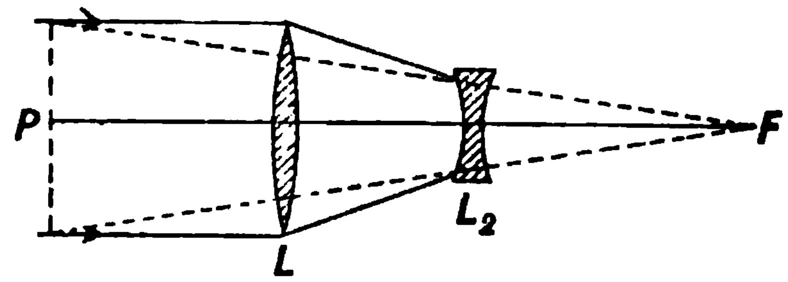


FIG. 572.

719. The Projection Lantern is used to throw an enlarged image $A'B'$ of more or less transparent objects on a screen. The object AB (Fig. 573) is also illuminated by a condenser C, consisting of two thick plano-convex lenses, with convex sides facing each other. The sources are usually the electric arc or the calcium light. The focusing lens L is generally of the

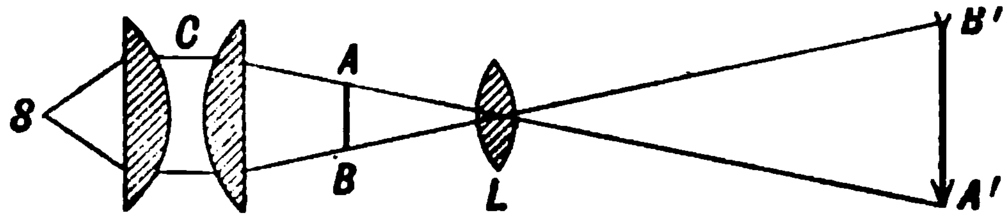


FIG. 573.

photographic doublet type, in order that an undistorted image may be formed on the screen. The object of the condenser is not only to illuminate the object, but also to enlarge the field beyond the limit which would otherwise be set by the cross-section of the focusing lens.

720. The Spectroscope (Fig. 574) is an instrument for analyzing complex radiations by prismatic dispersion (§672) or by the diffraction grating. In order to secure as complete separation of the colors as possible, or a "pure" spectrum, a narrow slit must be used as a source, so that the colored images of the slit will overlap as little as possible. The resolving power must be so great that the diffraction images of the slit or "lines" do not overlap, and this requires large apertures for the lenses and prism or grating (§707). The larger the dispersion the more complete the separation of the images. For given dispersion, the length of the spectrum is proportional to the focal length of the observing telescope, but this merely affects the scale of the spectrum, not the resolution of the lines or the clearness of detail.

The plan of the ordinary form of spectroscope is shown in Fig. 574. The essential parts are: A narrow slit, S; a collimating lens C, which converts the wedge of light from the slit into a parallel beam; a prism to disperse the colors; a telescope lens T, with which real images of the slit are produced at the focus of the eye-piece E. If light of an infinite number of colors is emitted by the source, the infinite number of partially overlapping images forms a *continuous* spectrum; if only a finite number of colors are emitted, there will be a finite number of slit images, giving a discontinuous or *line* spectrum. A homocentric pencil (cone) incident on a prism remains homocentric after transmission at the angle of minimum deviation; for all other angles of transmission, it becomes astigmatic, and no true image is produced.

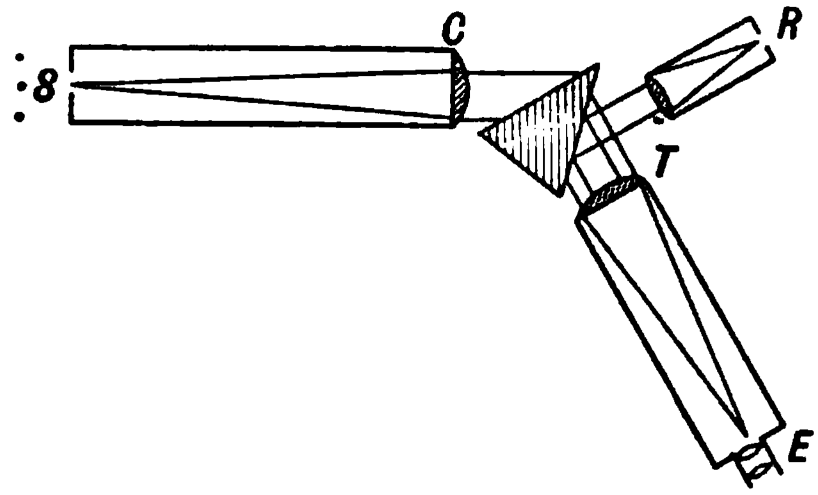


Fig. 574.

The condition of homocentricity cannot be fulfilled for all colors simultaneously except when the incident light is parallel, in which case light of each color emerges in a parallel beam. For this reason the collimator is necessary. This lens must be achromatic. So far as purity of spectrum is concerned, it is evidently unnecessary for the telescope lens to be achromatic, but it is usually corrected, in order that all the colors may be at once in the focus of the eye-piece. Positions in the spectrum may be referred to the image of a scale R reflected from the side of the prism.

This instrument is called a spectrograph when the telescope is replaced by a camera for photographing the spectrum, and a spectrometer when provided with a graduated circle for measuring the angular deviation of the light. The direct-vision spectroscope usually made in small sizes for pocket use, has a combination of

crown and flint prisms, as shown in Fig. 575. The mean deviation is zero, but there is some residual dispersion which gives a short spectrum (§676).

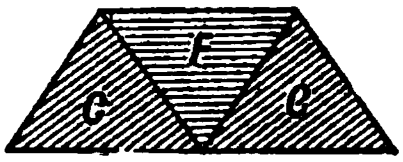


FIG. 575.

A plane diffraction grating may replace the prism of a spectroscope, or spectrometer. With the latter the angular deviations of the diffraction maxima may be measured and the wave-lengths determined by the relation deduced in §707.

721. The Concave Grating was Rowland's greatest contribution to spectroscopy. The lines are ruled at equal distances on the surface of a concave mirror of speculum metal which focuses as well as diffracts the light If R is the radius of curvature of the mirror (Fig. 576) and if the slit is at any point on the circumference of a circle having the radius as its diameter, it is found that the spectra of all orders are in focus, along the circumference of this same circle, so that no lenses are necessary. Usually the grating G s mounted at one end and the eye-piece, or camera, E at the other end of a beam R equal in length to the radius of the grating. This beam has a swivel truck under each end, which travels on tracks at right angles to each other, with the slit at the intersection S. The distance SE between the slit and the eye-piece are proportional to $a \sin \theta = n\lambda$ and therefore are proportional to the wave-lengths of the part of the spectrum in the field of view.

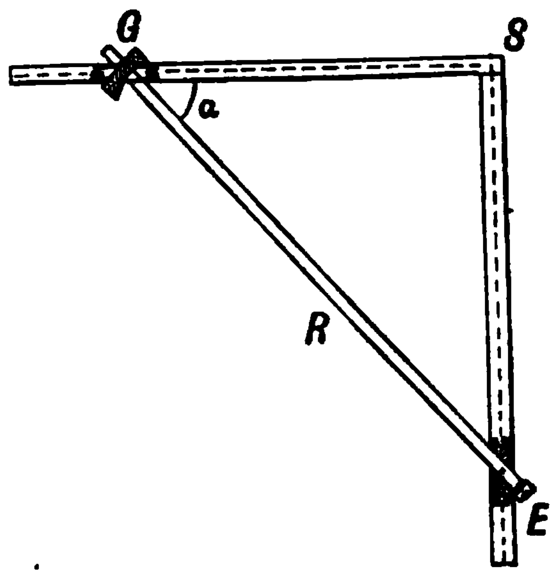


FIG. 576.

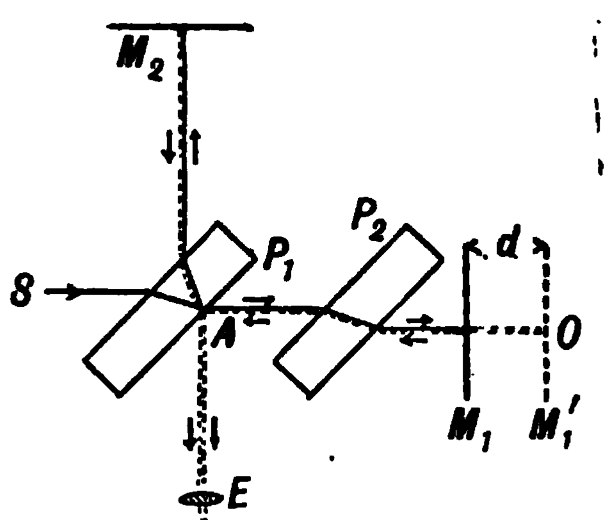


FIG. 577.

722. Michelson's Interferometer.—The surface of the glass plate P_1 (Fig. 577) is "half silvered," that is, the silver film is of such thickness that about one-half of the incident light is reflected. Light from the point S of an extended source falls on this surface at A and is in part transmitted to the mirror M_1, in part reflected to the mirror M_2. From these mirrors it will be reflected, retrace its course, and some will finally reach

the eye at E. If M_1 and M_2 are at the same optical distance from S, and if each is perpendicular to the rays that fall on it the light will appear to come from two exactly superimposed images of the source and there will be no interference. The plate P_2 is introduced merely to give the ray SAM_1 the same path in glass as the ray SAM_2, so that the optical and the geometrical paths will be the same. Now if monochromatic light of wave-length λ be used and if M_1 be displaced a distance $\lambda/4$, the waves that reach the eye will have traversed paths that differ by $\lambda/2$ and will destroy one another in the center of the field of view. A further displacement of $\lambda/4$ will restore the light. Thus by slowly displacing M_1 and counting the number of times, N, that the light reaches a maximum, the distance, d, through which M_1 has been displaced may be found from the relation

$$d = N\lambda/2.$$

Michelson has by this instrument measured the length of the standard meter of Paris in terms of wave-lengths of several of the spectral lines of cadmium, with an accuracy of about one part in ten millions. The wave-lengths of such lines are probably the most permanent and unchangeable standards of length which can be obtained. The interferometer has been used to make numerous other measurements of great accuracy. For example, the thickness of the thinnest water films has been found by means of it. If a film of thickness d be introduced in the path of one of the rays the change of optical path (§678) will be $2(n-1)d$. This distance being measured by the interferometer and n being known, d was deduced.

The above account is incomplete, since it was confined to what is observed in the center of the field of view, the mirrors being adjusted so that the image of M_1 in the plate P_1 is parallel to M_2. In this case the whole effect is the same as when light is reflected from a thin film with parallel sides (§695). Around a central point there are circular fringes which change in diameter as M_1 is moved. If M_2 and the image of M_1 are not quite parallel the effects will be similar to those due to a wedge-shaped film (§695). The fringes will still be approximately circular but not concentric, the centers being on a line perpendicular to the edge of the wedge. As M_1 moves the fringes will sweep past any point in the field of view and it is this succession of fringes that is usually counted in the use of the instrument.

The interferometer is also used to measure the difference of wave-length of two spectral lines which are so close together that they cannot be separated by a grating. Light from such a composite source produces two series of nearly coincident fringes. These, as M_1 is being displaced, coincide at equal intervals depending on the difference of wave-length and then produce maxima of visibility. The red lines of cadmium are apparently single lines.

723. Lummer-Brodhun Photometer.—A cube C is made of two right-angled prisms, as shown in Fig. 578. The hypothenuse surface of the one prism is plane, that of the other convex, with the vertex ground flat. These two surfaces are in close contact. The sources to be compared, S_1 and S_2, are mounted on an optical bench, over the center of which is the white screen W of paper or gypsum. The diffuse illumination from this screen is reflected from the mirrors M_1 and M_2 through the prism faces AB and CD. Light from S_1 and S_2 is transmitted without loss through the area of contact of the two prisms, and is totally reflected from the air film between those parts of the hypothenuse surfaces which are not in contact. If a telescope is focused on the region of contact through the side CD, light will enter it from S_1 by transmission and from S_2 by reflection. The field will be uniformly illuminated if the two sides of the screen W are equally illuminated; otherwise the area of contact will appear brighter or darker than the surrounding part of the field.

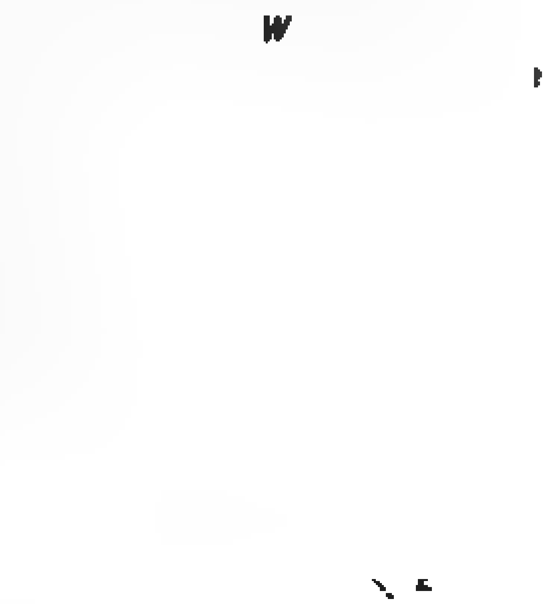


Fig. 578.

724. Standards of Luminosity.—For accurate and easily reproducible comparisons of photometric measurements constant and easily attainable standards are necessary. So far no absolutely reliable standard source has been found. For ordinary purposes the British standard candle is used. These are made of sperm, weigh six to the pound, and are normally supposed to burn 120 grains per hour. In actual practice there are great deviations from uniformity.

Other standards in use are the Methven screen and the Hefner-Alteneck lamp. The former is a gas flame from an Argand burner, the light from which passes through a rectangular opening of definite size. The latter is the flame of a lamp burning amyl acetate. In both, the flame should be kept at a definite and constant height. The light from these sources is constant within a few per cent.

EMISSION OF RADIANT ENERGY.

725. Analysis of Radiation.—The methods by which radiation may be analyzed by the dispersion of colors, or waves of different

length, have already been described (§720). The radiation from all known sources is complex—that is to say, it contains waves of more than one frequency of vibration. The principal types of emission spectra may be observed side by side if the image of a long electric arc is focused on a slit beyond which a prism and lens are placed so that a large spectrum is thrown on the screen. The light coming from the positive carbon forms a brilliant continuous spectrum including all the colors. Next to it is the discontinuous spectrum of the arc proper, the luminous flame, which contains the vapors of carbon, various compounds of carbon, and any metals that may be present as impurities in the electrodes. This spectrum consists of a number of narrow lines due to the metals present, and several groups of bands, each composed of a large number of fine lines so spaced as to produce the effect of the shading of fluted columns in a line drawing—hence they are often referred to as fluted bands (see Fig. 582). These appear to be due to the vapors of carbon or the compounds of carbon. The bands are especially strong in the violet, and this gives the arc its characteristic violet color. The violet bands appear to be mainly due to cyanogen or some other compound of carbon with nitrogen, as they are very weak if the arc is deprived of nitrogen. Next to this is the spectrum of the negative electrode, which is at a much lower temperature than the positive, in which there is very little blue or violet.

As illustrated by the arc spectrum, there are two general types of *emission* spectra, the *continuous* and *discontinuous*, and the latter in turn may be divided into *line* and *band* spectra. §§731, 732.

726. Invisible Radiations.—It was found by William Herschel in 1800 that if a sensitive thermometer is placed in any part of the spectrum of the sun it will show a rise of temperature, this effect increasing in going from violet to red. It does not, however, cease abruptly at the boundary of the visible spectrum, but increases to some distance beyond it, and then gradually diminishes, the observed limit in any case depending on the sensitiveness of the thermometer. Evidently there is radiation which is less refracted than the red, and which by analogy we may conclude has waves of greater length than those of red light. It was shown by Herschel that this "radiant heat" is subject to

the same laws of reflection and refraction as light, but their identity was not generally accepted until nearly fifty years later, when it was shown that the invisible radiation is capable of producing interference effects, and that it is likewise capable of dispersion and polarization. As the ideas in regard to the nature of heat crystallized it was seen that heat can be associated only with matter, but that the energy of this heat may be partly transformed into the energy of ether waves, and this, if absorbed by matter, will again appear as heat. It thus becomes clear how invisible radiations from a hot body can pass through an ice lens without melting it, and set fire to a piece of paper at the focus. The name *infra-red* is applied to these long-wave radiations.

The existence of *ultra-violet* radiations in the solar spectrum, with waves shorter than those of violet, is shown by means of the chemical effect produced on chloride of silver. Photographic films are very sensitive to the ultra-violet radiation, which is especially active in its chemical effects. It also excites strong fluorescence (§750) in many substances. If a strip of paper moistened in acidulated sulphate of quinine solution is held in the arc spectrum the excited fluorescent light shows the existence of ultra-violet radiation in the spectra of both the positive carbon and the arc proper.

The non-visibility of the infra-red and ultra-violet radiations is due merely to the limitations of the eye. The eye will "resonate" to vibrations between certain limits of frequency, the photographic film or fluorescent screen to certain others; but if the receiving surface is blackened the energy of waves of all frequencies is almost completely absorbed, and by the amount of heat developed we may determine the amount of energy in any part of the spectrum.

727. Methods of Detecting Invisible Radiation.—Photography is a thoroughly satisfactory method of detecting even the shortest ultra-violet radiations so far discovered. This method cannot be used however, at the opposite end of the spectrum, as it seems impossible to make any photographic film which is sensitive to the infra-red—it is difficult to make one which will even reach the limit of the visible red For this reason other less satisfactory methods must be employed, which are usually based on the heating effects produced. For one of the earliest instruments for the detection of infra-red radiation, the thermopile, as well as other more recent modifications, see §§333, 482.

728. Continuous Spectra.—It is a familiar experience that as the temperature of a body rises it first reaches a dull red heat, then yellow, and finally a dazzling white. Conversely, if the spectrum of the positive electrode of an arc light is thrown on a screen, and if the current is suddenly cut off, it will be observed that, as the carbon cools, violet, blue, green, and yellow disappear in succession, and finally the red. If a sensitive thermopile is placed far in the infra-red it will be found that sensible radiation is still emitted long after the luminosity has disappeared.

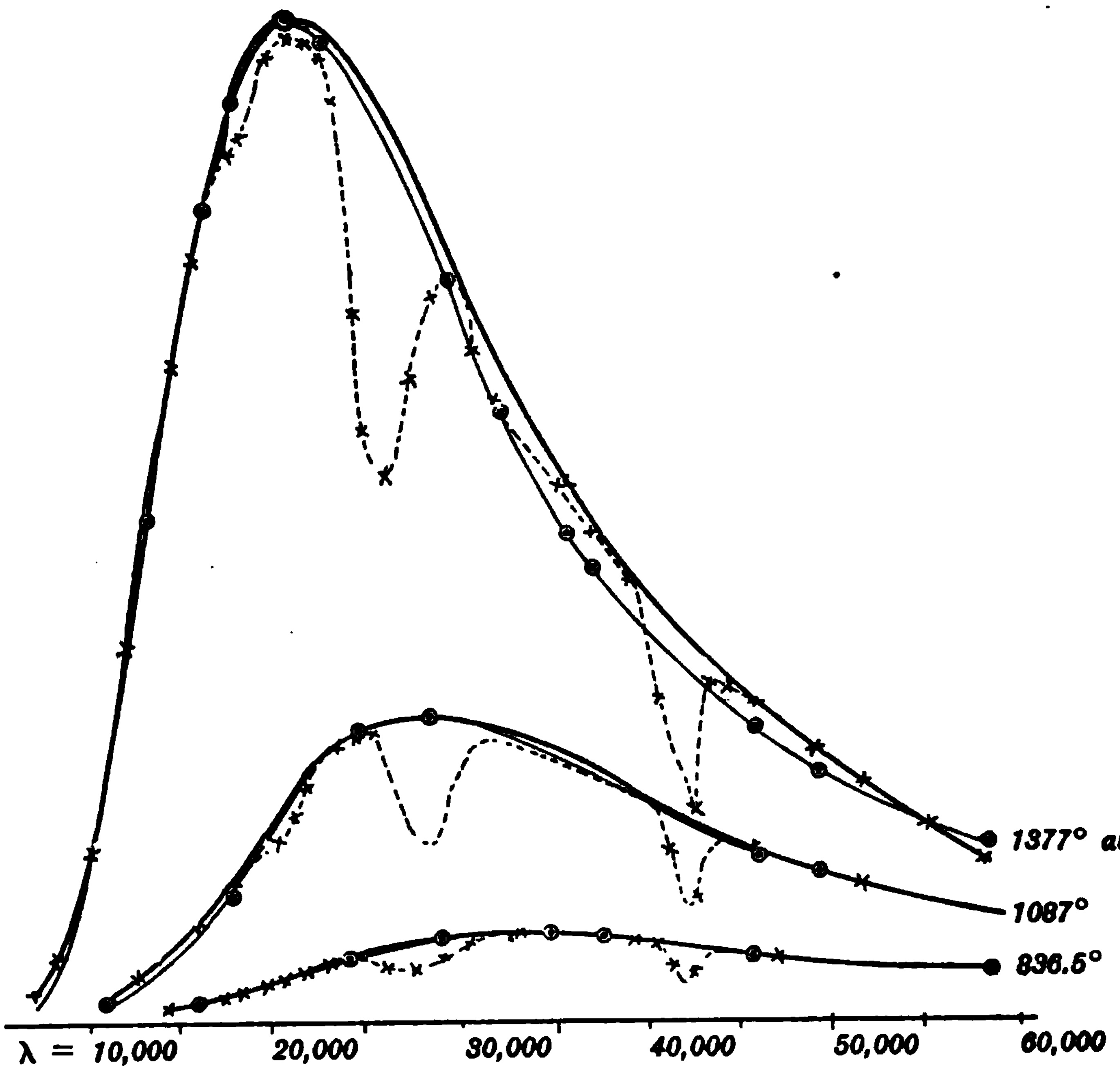


Fig. 579.

Draper (1847) found that all bodies begin to glow at about the same temperature. The actual temperature in any case depends somewhat on the sensitiveness of the eye, but is not far from 400°. *Draper's law* is approximately true for all colors and temperatures—that is to say, all solids begin to radiate red, or yellow, or violet, or any particular "heat color" in the infra-red at the same temperature. The spectral distribution of energy

may be shown by plotting a curve with wave lengths as abscissæ and ordinates proportional to the galvanometer deflections observed as the thermopile passes through the spectrum. Fig. 579 shows a series of such curves for temperatures ranging from 836° to 1377° absolute, the source consisting of a strip of blackened metal electrically heated. The depressions in the curves are due to absorption by carbon dioxide and water vapor. The general character is the same for all solids, but differences in the ordinates may arise from differences in the state of the surface, whether black, or rough, or polished, etc. Investigation shows that there is a very definite relation between the absolute temperature of the source if it is black, or approximately so, and the wave length corresponding to the maximum ordinate of the energy curve, such that $\lambda_m T =$ Constant (see §729). This constant varies slightly from 2814 at 621.2° to 2928 at 1646°, with a mean value of 2879. The unit of wave length is the micron, μ, or .001 mm.

The total energy emitted by an incandescent source is proportional to the area included between the energy curve and the axis of X. That part of the energy which produces luminosity is included between the limits of the visible spectrum. The luminous efficiency is proportional to the ratio between these two areas. Fig. 580 illustrates the relative luminous efficiencies of the positive pole of an arc light, of an incandescent light, and of a piece of red hot carbon. The luminous energy is represented by the shaded part of the area between the curve and the λ-axis. Evidently the luminous efficiency rises very rapidly with the temperature, so that a small increase in the current through an incandescent lamp will greatly increase its brightness, and conversely. The luminous efficiency of the arc is about 10 per cent.; of incandescent lamps, 3 to 5 per cent.; of gas and candle flames 2 or 3 per cent. Luminous vapors, which radiate "selectively," are usually more efficient than solids.

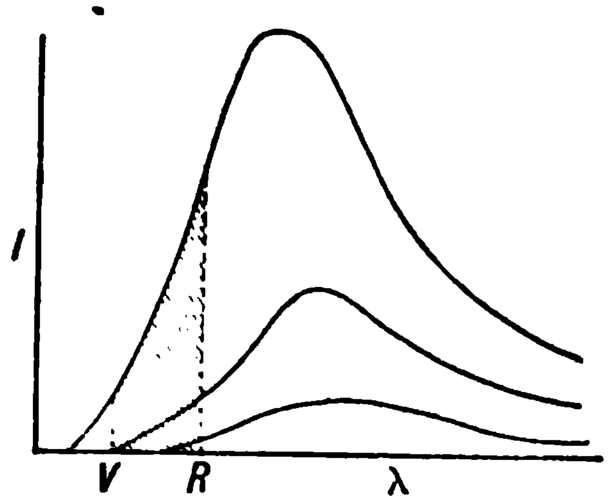


FIG. 580.

729. Law of Radiation.—It may be assumed that ether waves are set up by agitation of the electrons, or ions in the case of the long waves, within the

molecules of matter, and that the frequencies of vibration of these electrons or ions is dependent upon the kinetic energy of the molecules (§262). In solids the molecules are so close that there can be little chance for the electrons to vibrate without constraint with their natural periods. Owing to frequent collisions, a wide range of velocities and vibration frequencies must exist. The forced ether vibrations will have periods corresponding to those of the electrons. The greater part of the radiant energy will be due to the large number of molecules which have velocities in the neighborhood of the mean velocity. There will be relatively few molecules which will have extreme velocities, either large or small, and therefore the longest and shortest ether waves will have relatively a small amount of energy. As the temperature rises there will be a general increase of kinetic energy, many molecules moving faster but none slower than before, so that both the maximum energy and the maximum rate of vibration of the excited ether waves will move toward the violet end of the spectrum. The source will rise to red and finally to white heat. It is thus evident that a spectral intensity curve must be a sort of probability curve based on the distribution of velocities of the molecules with its ordinates exaggerated on the side of the violet, as illustrated in the curves of Fig. 579.

By such reasoning Wien theoretically deduced the relation $\lambda_m T$ = constant, known as Wien's law. Later Planck established a general relation between the intensity of radiation corresponding to a given wave-length λ, and the absolute temperature of the source, as follows:

$$I_\lambda = \frac{C}{\lambda^5}\left(e^{\frac{c}{\lambda T}} - 1\right)^{-1}$$

where e is the base of the natural system of logarithms, T the absolute temperature, and C and c constants. This relation holds within wide limits for bodies which are black or approximately so. For such bodies the value of $c = 1.4598$, $C = 3.7179 \times 10^{-5}$. The law $\lambda_m T$ = constant $= c/5$ may be deduced by differentiating the above expression for a maximum value of I_λ, and Stefan's law by integrating the intensity over the whole spectrum (see §337).

730. Discontinuous Spectra were noted by a number of observers during the first half of the nineteenth century, and it was at least dimly realized that in some cases the appearance of the spectrum is characteristic of the substances emitting the radiation. There are two types of discontinuous spectra, known respectively as *line* and *band* spectra.

In 1860 Kirchhoff and Bunsen established definitely the law that all gases and vapors give discontinuous spectra, and that these spectra are perfectly characteristic of the substance. They discovered rubidium and cæsium by the application of this principle,

which has been fruitful in the discovery of other new elements, notably helium and the rare atmospheric gases in recent times.

At first only the spectra of flames were studied, but later it was found that the electric spark between metallic terminals gives lines due both to the electrodes and to the surrounding atmosphere, and that if the electric discharge passes through a gas in a partially exhausted tube (vacuum tube) the luminosity is confined to the gas, and the metallic lines disappear. Only in exceptional cases is it possible to make a gas luminous except by the electric discharge.

It would seem reasonable to imagine that an electron attached to an atom has its own definite rate or rates of vibration, just as a tuning fork has one definite period and a piano wire several. In a solid or liquid constraints and collisions may produce forced vibrations covering a wide range of periods, but in gases or vapors, where collisions must be comparatively few, there is a preponderance of free vibrations. If an electron has one free period of vibration, say that corresponding to the color of luminous sodium vapor, there will be but one image of the slit if the D lines were single—yellow in the case assumed. If there are a number of coexisting vibrations of different periods there will be an equal number of spectral lines.

731. Line Spectra are given by the metals and salts of the sodium and calcium groups in the Bunsen flame, and also by a number of other metals if spray from solutions of their salts, or ions caused by the electric spark, are passed into the flame. The spectrum of the electric spark or arc between electrodes composed of or coated with any metals or their salts contains many more lines than that of the flame. The number may be very large, ranging from a dozen or so in the case of the alkali metals to many thousands, in the cases of iron and uranium.

It seems that no two elements have any common lines, but the spectrum of a given element may show differences in the number and the appearance of the lines according to the nature of the source, whether flame, arc, or spark. The middle part of Fig. 581 shows the arc spectrum of iron, the lowest part is the spark spectrum of the same metal, the highest part is the solar spectrum, which shows many absorption lines coinciding with the emission lines of iron. All salts of the same metal give the same line

spectrum, although in some cases they give band spectra as well, which may be characteristic of the salt (see next section). The lines of the non-metallic components do not appear with those of the metal except in rare cases. Intense electric discharges through a non-metallic gas at ordinary pressures or in vacuum tubes give line or band spectra.

Fig. 581.

732. Band Spectra are usually composed of fine lines, as shown in Fig. 582 (the spectrum of part of the carbon arc, supposed to be due to cyanogen). The light of the green cone in a Bunsen flame gives a very similar spectrum, due to carbon or its compounds. The salts of the calcium group of metals have flame spectra containing both lines and bands. All salts of calcium, for example, give the same flame spectrum under ordinary conditions, but if calcium chloride, for instance, is placed in a flame supplied with hydrochloric acid an entirely different band spectrum is

Fig. 582.

produced, and still another if the flame is supplied with calcium bromide and an excess of hydrobromic acid. The inference is that in these cases the bands represent the characteristic spectra of the compounds, and that the spectrum observed under ordinary conditions is that of the oxide, due to reaction with atmospheric oxygen.

Nitrogen gives a band spectrum very similar to that of cyanogen if a feeble discharge passes through it, but an entirely different line spectrum if the discharge is very intense. (Fig. 583.)

Nitric oxide gives a characteristic band spectrum in the ultra-violet similar to that of nitrogen. All the compounds of mercury with chlorine, bromine, or iodine, give characteristic band spectra with feeble discharges, and the line spectrum of mercury with strong discharges. The same is true in a number of other cases. All these facts are consistent with the view that band spectra are characteristic of the molecular state of either elements or compounds. Intense discharges, by dissociating the molecule, will produce line spectra, characteristic of the atomic or, rather, "ionic" state. The salts of the alkali metals are so easily dis-

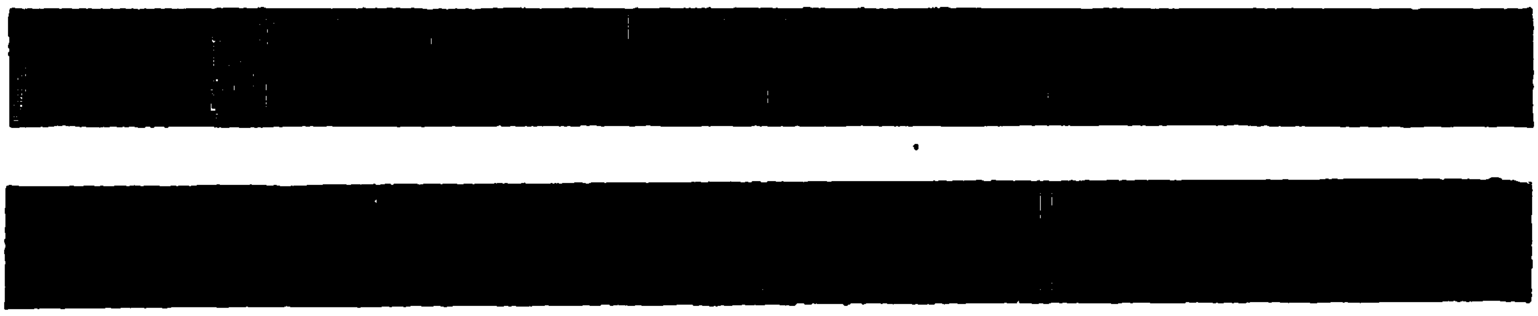


Fig. 583.

sociated that they give only line spectra in the flame. It is possible, however, by other modes of excitation to produce band spectra of these elements (see §751).

733. Limits of the Spectrum.—The very short ultra-violet waves are absorbed by all gases except hydrogen, and by most lenses and prisms. Working with fluorite lenses and prisms or a grating in a vacuum, Schumann and Lyman have reached a wave-length of about .00006 mm. The ordinarily used unit of wave-length is the Ångström unit, equal to one ten-millionth of a millimeter. This is sometimes called a tenth-meter. Another unit frequently used is the micron, $\mu = 0.001$ mm. Expressed in these units, some wave-lengths are given below. Most substances are opaque to very long waves, and some of the longest waves mentioned were obtained by the method of selective reflection described in §778, the wave-length being then measured by a coarse grating.

It has been proved that X-Rays (§553) and gamma rays (§575) are similar to light waves, but very much shorter. Their lengths differ, but the order of their magnitude is given in the table below.

	Ångström Units	μ
Gamma rays	0.1	0.00001
X-rays	1	0.0001
Shortest ultra-violet waves	600	.06
Shortest visible waves (violet), about	3,800	0.38
Violet, about	4,000	0.4
Blue	4,500	0.45
Green	5,200	0.52
Yellow	5,700	0.57
Red	6,500	0.65
Longest visible waves (red)	7,500	0.75
Longest waves in solar spectrum, more than	53,000	5.3
Longest waves transmitted by fluorite	95,000	9.5
Longest waves by selective reflection from rock salt	500,000	50.0
By reflection from potassium chloride	612,000	61.2
Longest waves from mercury lamp	3,140,000	314
Shortest electric waves	40,000,000 = 4 mm.	4000

734. General Absorption.—When radiation falls on matter a portion is reflected, another absorbed, and if the substance is transparent or very thin a part is transmitted. Black substances, such as lampblack and copper oxide, reflect and transmit very little, the absorption being almost complete. Most substances black to visible radiation are also black to the ultraviolet and infra-red waves, but there may be exceptions—for example, a sheet of hard black rubber is opaque to visible radiation, but transparent to waves beyond the red. Substances like that last mentioned, which absorb certain radiations and transmit others, are said to exercise selective absorption.

735. Selective Absorption is characteristic of most substances. Familiar examples are red glass, which transmits red and some infra-red, but no other visible colors; blue cobalt glass, which transmits blue and violet and a little red and green in narrow regions; green, which transmits almost all the colors, but a larger proportion of green; chlorophyll solution, potassium permanganate, the aniline colors, and solutions of the rare earths, didymium, etc. In most cases the absorption bands are wide and diffuse; in the case of the rare earths they are almost as narrow as spectral lines, so that the solutions appear almost colorless, no large amount of any one color being absorbed; the vapors of iodine, nitrogen peroxide, and some other substances have fluted absorption bands, grouped somewhat like the lines

in the nitrogen bands. Many substances such as glass, quartz and rock salt are very transparent within wide limits, beyond which they are completely opaque.

Glass is opaque to waves shorter than 3500 Ångström units, and longer than about 30,000 Ångström units. Quartz is transparent between the wave-lengths 1800 and 70,000, and for some longer waves; rock salt s transparent between 1800 and 180,000, and fluorite, one of the most transparent substances, will transmit ultra-violet waves from about $\lambda = 1000$ to $\lambda = 95{,}000$.

736. Kirchhoff's Law.—If the fraction A of the radiation of a given wave-length incident on a body is absorbed, A is said to be its absorbing power for that color. The *emissivity* of a radiating body is the amount of energy radiated per second from each unit of surface. Kirchhoff showed by the theory of exchanges (§335) that the emissive and absorptive powers of all bodies at the same temperature for a given color are proportional when the radiation is a pure temperature effect.

737. Origin of the Fraunhofer Lines. A general account of these lines has been given in §673.

Kirchhoff, noting that there were coincidences between many of the Fraunhofer lines and emission lines, explained them as the result of absorption by vapors in the sun's atmosphere of waves which these vapors emit themselves. Stokes independently suggested that the coincidence of the yellow sodium lines with the D lines indicated that the sodium atoms must absorb waves of the same frequency as those emitted by them, the effect being similar to resonance phenomena in sound. This *reversal* of the sodium lines is easily secured by igniting a small piece of metallic sodium in a metal spoon before a slit illuminated with the electric arc, the light then passing through a prism and a lens which focuses it on a screen. If a large quantity of sodium vapor is present in an arc the phenomenon of *self-reversal* is shown in the spectrum. The bright lines are very broad and intense, with a narrow dark line in the middle of each, due to absorption by the cooler sodium vapor in the outer portion of the arc.

738. Luminescence.—In all cases where radiation is purely a temperature effect Kirchhoff's law appears to hold. In many cases, such as those of fluorescence and phosphorescence (§§750. 751), in which the absorption of waves of certain lengths causes

the emission of waves of a different length, this is not true; nor is it generally true of luminous gases and vapors, where the luminosity appears to be due to electrical or chemical causes. In no known case do gases or vapors have absorption lines corresponding to all the emission lines. The name *luminescence* has been applied to the various kinds of radiation not directly due to high temperature and not conforming to Kirchhoff's law.

739. Solar Spectrum.—The wave-lengths of many thousands of the Fraunhofer lines were determined by Rowland. A large number were found to coincide with the emission lines of known elements, so that it seems certain that about forty of these elements exist in the sun. (See Fig. 581, which shows the coincidence of many of these absorption lines with the emission lines of iron). The chromosphere, or gaseous solar atmosphere, the prominences or flames of incandescent hydrogen and other gases rising out of it, and the corona, or nebulous outer envelope, give bright line spectra, which may be seen during a total eclipse, when the brighter light from the photosphere does not mask them. The rare gas helium was known to exist in the sun before it was found on the earth, on account of the bright yellow line due to it observed in the spectrum of the prominences.

The ultra-violet region of the solar spectrum does not extend beyond a wave-length of about 3000 Ångström units. Without doubt shorter waves are emitted, but they are absorbed by the earth's atmosphere, which is opaque to all very short waves. The atmosphere also exercises general and selective absorption in the visible region. Oxygen and water vapor give rise to the terrestrial lines and bands known as the Fraunhofer lines A, α and B, and there is more or less general absorption due to these and other constituents of the earth's atmosphere.

Wave-lengths of Fraunhofer Lines

A	7594–7621	O	Red
B	6870	O	Red
D_1	5896.15	Na	Orange
D_2	5890.18	Na	Orange
E_2	5269.72	Fe	Green
F	4861.50	H	Blue
g	4226.89	Ca	Violet
H	3968.62	Ca	Violet
K	3933.81	Ca	Violet

The infra-red region of the solar spectrum has been investigated by Langley with the bolometer, and found to extend beyond a wave-length of 53,000 Angström units. Broad absorption bands are found, some of which coincide with those due to water vapor and carbon dioxide, besides many narrow lines and bands of unknown origin. A large proportion of the solar radiation, particularly in the neighborhood of the shorter waves, is absorbed by the earth's atmosphere, and this must greatly influence climatic conditions.

740. Spectra of Planets, Stars, Comets, and Nebulæ.—The planets and the moon give spectra similar to that of the sun, as might be expected, but modified by general and selective absorption in the cases of the planets which have an atmosphere. Most stars have characteristic absorption spectra resembling that of the sun, which shows the universal distribution of many of the common elements. In addition there are frequently lines due to unknown elements. Nebulæ give bright line spectra, some of the lines being due to hydrogen and helium, while others have not yet been identified. The spectrum of comets consists mostly of the characteristic hydrocarbon bands similar to those given by the green cone of the Bunsen flame. It seems evident in the cases of nebulæ and comets that the radiation is an example of luminescence, or luminosity due to other causes than high temperature, because these bodies appear to consist of masses of highly attenuated gases, or small bodies, and it is inconceivable that their temperature can permanently remain much higher than that of the surrounding space.

741. Application of Doppler's Principle.—If a star is approaching or receding from the earth, the effect will be to shorten or lengthen each wave reaching the earth (§596). Each line will be displaced toward the violet if the star is approaching, toward the red if it is receding. By measuring such displacements on photographs of stellar spectra the velocities of stars in the line of sight may be determined with an error of less than one kilometer per second. Most of the stars which have been investigated have velocities with respect to the sun of between one and one hundred kilometers per second. It is found that a majority of the stars on one side of the heavens have a general relative motion toward the sun, those on the opposite side away from the sun. The inference is that the solar system is itself moving through the universe in the former direction.

Fig. 584 is the spectrum of χ Draconis, with comparison spectra of hydrogen above and below, showing the Doppler effect on the hydrogen absorption lines.

FIG. 584.

EFFECTS DUE TO ABSORPTION

742. Color of Natural Objects.—The colors seen in the spectra produced by dispersion or by interference are pure. This is not the case with the colors of natural objects, which as a rule are due to selective absorption of certain colors of the incident light, the other colors being diffusely reflected in different proportions. If a colored object, such as a red rose, is placed in different parts of a spectrum, it will appear a brilliant red in the red and almost black in other parts. This shows that the greater part of all colors except red is absorbed; not all, however, for it will be noticed that in every part of the spectrum there is some reflection of the incident color. Since the resultant of the combination of all colors is white, it may thus be proved that from all colored objects some white light is reflected, in addition to the characteristic color.

743. Body Color.—In most cases it is observed that bodies having a certain color by reflected light have the same color by transmitted light. This suggests that the color diffusely reflected is due to components of the incident white light which have penetrated more or less into the medium before being scattered, the other colors being lost by absorption. The white light reflected is probably due both to reflection at the surface and to the recombination of the various colors which escape complete absorption. As a crude illustration of body color, if light falls on a piece of red glass a white image of the source will be reflected from the front surface and a red image from the rear surface.

Colors are said to be more or less *saturated* according to the proportion of white light with which they are diluted. The pure spectral colors are said to be completely saturated. The pro-

portion of white light scattered is increased by any process which increases the reflecting surface. For example, crystals of copper sulphate will appear lighter and lighter as they are crushed into smaller fragments, and become almost white when reduced to a fine powder. The white light reflected from the numerous surfaces then completely masks the small portion which is selectively transmitted. Similarly, transparent substances such as glass are white when in powdered form.

744. Dichromatism.—Some substances when examined by light transmitted through thick layers appear to be of different color from that observed by reflection or by transmission through a thin layer. A thin layer of chlorophyll is green by transmitted light, while a thick layer is red. This is explained by the fact that the absorptive power or the fraction of the incident light absorbed by a layer of unit thickness, is different for the two colors. While the incident green light is more intense than the red, and remains so after transmission through a thin layer, it is more rapidly cut down by absorption, so that after passing through a thick layer the red predominates. This effect is called dichromatism.

745. Surface Color.—Some substances appear of different colors by reflected and by transmitted light. Such is the case with thin films of metal and of the solid aniline colors. Gold is always yellow by reflected light, but a sheet of gold leaf thin enough to permit transmission appears green by the transmitted light. The light reflected from these substances is complementary to that transmitted. In such cases selective action seems to take place at the surface, some colors being directly reflected, others being absorbed by a thick layer, or transmitted through a thin film. Bodies exhibiting surface color retain that color when finely powdered.

746. Colors of Sky and Clouds.—Since light can reach the eye only directly from the source or by reflection from material objects, it is evident that, since the sky is not perfectly black, it must contain matter in suspension. Some have supposed that air itself may have a characteristic color, as is shown by great thicknesses of glass or of water, but it is probable that the blue color of the sky is due to selective scattering by small suspended particles of dust, water, etc. It is to be expected

that such small particles should reflect a larger proportion of short waves than of long ones. The term scattering is used, because it seems evident that this is not a case of ordinary reflection like that from a mirror of finite size. There is an analogy in the case of sound waves; long waves pass around obstacles without deviation from their general direction, while shorter waves may be reflected. Since the shorter waves of light are scattered, the transmitted light will consist mostly of the longer waves. This accounts for the brilliant reds, oranges, and greens often observed in the western sky at sunset. The light transmitted almost tangentially through the atmosphere has been deprived of the shorter waves, which cause a blue sky for those more immediately under the sun. These effects are intensified by the presence of a large number of dust particles in the lower levels of the atmosphere. After the great eruption of the volcano Krakatoa in 1883 fine volcanic dust pervaded the atmosphere of the whole earth and the sunsets were especially brilliant. For the same reason lights look red when seen through smoke or fog, or through water made slightly turbid by the addition of a small quantity of milk or shellac solution. This effect is beautifully illustrated by passing a beam of light through a jet of steam issuing from a small nozzle into a stream of air previously dried by forcing it through sulphuric acid. The size of the water drops is controlled by changing the vapor pressure in the atmosphere in which the drops are formed, lower vapor pressure promoting evaporation and thus reducing the size of the drops. The colors seen by transmitted and by scattered light are complementary, the shorter waves being scattered and the longer ones transmitted.

747. Color Sensation.—The perception of a given color by the eye does not necessarily prove that the stimulus is of the corresponding wave-length. It may be the resultant effect of several different colors. For example, if the light from the red of a spectrum and from a region intermediate between the blue and the green be superimposed the resultant sensation is white, which the eye cannot distinguish from the white due to a mixture of all the colors. A similar effect is produced by the combination of violet and yellow-green. Two colors which together give the sensation of white are said to be *complementary*. It is found,

further, that spectral red and green combined excite a sensation of yellow, while green and violet produce blue. All possible colors may be produced by combining red, green, and violet. According to the theory of Thomas Young, these are to be regarded as the three primary color sensations. The cones in the retina are supposed to respond or "resonate" most actively to frequencies of vibration corresponding to these colors, and all color sensations depend on the proportions of the incident energy belonging to these frequencies.

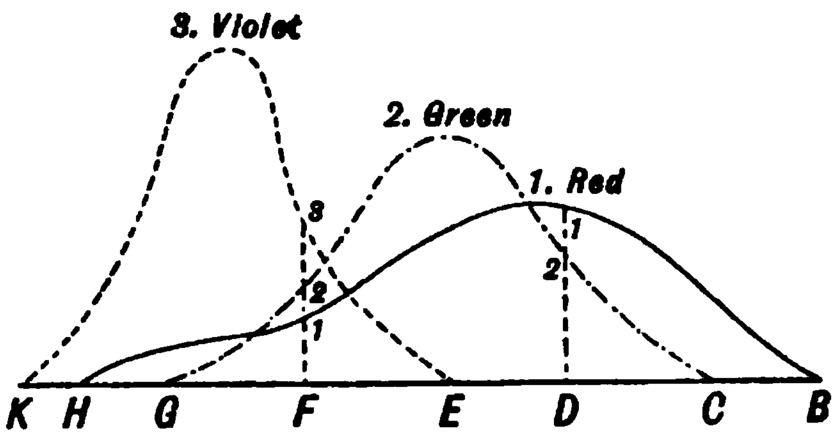


FIG. 585.

The phenomena of color sensation may be explained by assuming that in the normal eye there are three sets of nerves, one stimulated most actively by red light, another by green, and a third by violet, but each responding also more or less to waves of other frequencies as well. To indicate these effects Koenig constructed three curves (Fig. 585) on an axis representing the length of the normal spectrum from the Fraunhofer line K in the violet to B in the red. The ordinates at each point represent the degree of excitation of the three sets of nerves respectively by light of the frequency corresponding to that point in the spectrum. The maximum sensibility of the "red" nerves is in the orange-red, that of the "green" nerves in the green, and that of the "violet" nerves in the blue-violet. The first set of nerves is also excited more or less by all colors between H and B; the green by all colors between G and C, and the violet by all colors between K and E. The color of sodium light (D) is caused by the superposition of two sensations, red and green, proportional respectively to the ordinates $D1$ and $D2$. The color of the blue line of hydrogen (F) is due to a combination of red, green, and violet sensations proportional to the ordinates $F1$, $F2$, and $F3$. In the case of color-blind persons one or more sets of nerves are missing—usually the red. To such persons, for example, sodium light would appear green. Two colors, such as red and blue-green, are complementary, when, acting jointly, they excite all three sets of nerves in the proper proportion to produce the sensation of white—or in the same proportions that they are excited by ordinary white light.

748. Pigment Colors.—The effect of mixing pigments is quite different from that of mixing spectral colors. For example, blue paint absorbs nearly all the incident light except the blue and some green; yellow paint absorbs nearly all except yellow and some green. If, therefore, white light is incident on a mixture of the two pigments green is the only color which escapes absorption by one or the other, therefore a mixture of blue and yellow paints produces a green paint. In such cases the apparent color may vary with the kind of illumination. Blue pigments usually appear green by candle light, because there is a very small proportion of blue in the incident light, and so green predominates in the scattered light.

749. Chemical and Molecular Effects.—Light may cause chemical combination, as when it acts on a mixture of hydrogen and chlorine, or dissociation, as when it acts on the silver salts in a photographic plate. By its action on the chlorophyll of plants, light decomposes the carbon dioxide absorbed from the atmosphere, releasing the oxygen and causing the carbon to be assimilated. It may cause molecular transformations, as when it alters amorphous to crystalline selenium, or changes the electric resistance of the latter form. It also changes white phosphorus to red. These effects are not due to the heating effect of the absorbed radiation, because an equivalent rise of temperature will not cause them, but they seem rather to depend on the vibratory character of the light waves. As a rule the shorter waves are the most effective in producing such results.

Another effect due to light, especially to the ultra-violet waves, is that it will cause the discharge of electricity from certain metals (§565).

750. Fluorescence.—There are substances which when stimulated by the absorption of waves of certain lengths will emit waves of different lengths. For example, a piece of paper moistened with sulphate of quinine solution and held in the ultra-violet portion of the solar spectrum will emit a brilliant opalescent blue light. To this phenomenon Stokes gave the name of fluorescence, because it was observed in fluorspar. He explained it as the result of the absorption of incident waves which by a modified resonance action caused a reëmission of longer waves. Similar effects are observed in coal oil, fluorescein, eosin, uranin, and other organic compounds; in uranium glass, which emits a yellowish-green light; in esculin, which emits blue light, and in chlorophyll, which emits red light; and also in a much smaller degree in iodine, wood, paper, and many other substances.

751. Phosphorescence.—There is a large class of substances, of which calcium, strontium, and barium sulphides are familiar examples, which after exposure to light show effects which are similar to fluorescence, but which continue visible long after the exciting radiation ceases to act. This is called phosphorescence. The only definite distinction between fluorescence and phosphorescence is that the latter persists for a longer time. Many substances which phosphoresce very feebly at ordinary temperatures may be made to glow brilliantly at the temperature of liquid air. As examples, gelatin, horn, egg shells, and paper may be mentioned.

Some metallic vapors, such as those of the sodium and calcium group, fluoresce brilliantly under the action of light or cathode rays. The light shows the characteristic spectral lines and bands of the metal. Certain organic vapors, such as anthracene, fluoresce when light falls on them. Nitrogen, oxygen, and some other gases will under certain conditions phosphoresce brightly for several seconds after an electric discharge has passed through them in a vacuum tube (§550).

DOUBLE REFRACTION AND POLARIZATION

752. Double Refraction.—Some crystals, such as those of rock salt and fluorite, resemble isotropic solids, such as glass, in the respect that their physical properties are alike in all directions. In general, however, this is not the case; such properties as elasticity and heat conduction, as well as optical properties, differ in different directions in the crystal. In such crystals as quartz and calcite there is an *axis of symmetry*, the crystallographic axis, and the physical properties are the same in all directions in any equatorial plane, but different from those in the direction of the axis. Iceland spar, or calcite, is a rhombohedral crystal, each face being a parallelogram with two acute angles of 78° 5′ and two obtuse angles of 101° 55′. Two solid angles of the crystal are formed by the junction of the obtuse angles of three faces. Any line equally inclined to the faces of one of these solid angles is a crystallographic axis. An object seen through Iceland spar appears double, unless viewed in the direction of the axis. No such effect is observed in the case of isometric crystals. This

phenomenon is called *double refraction.* When the waves travel in the crystal in the direction of the crystallographic axis there is no double refraction; hence ***any line in the crystal parallel to the axis*** is called an ***optic axis.***

If a ray *r* of ordinary light is incident normally on any face of a doubly-refracting crystal one ray *o* is transmitted without deviation; and if the incidence is oblique (Fig. 586) this ray is deviated,

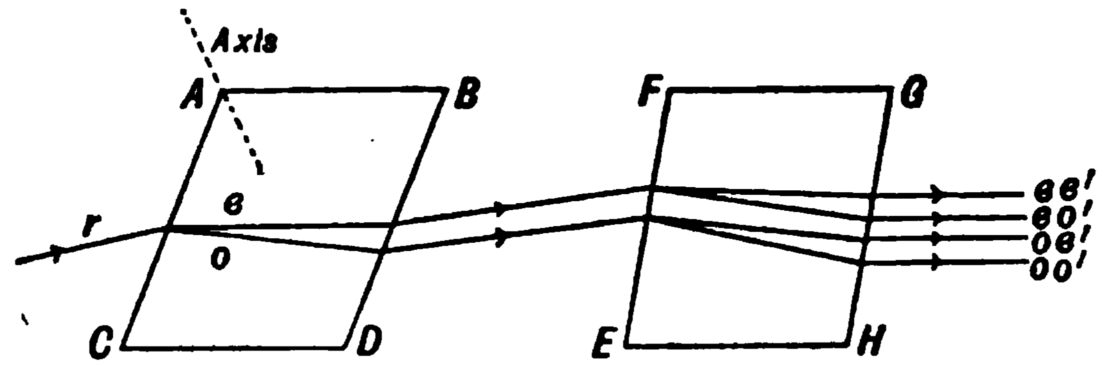


Fig. 586.

with an index of refraction which is independent of the angle of incidence. The other ray *e* is deviated in all cases, unless it travels along an optic axis, and the index of refraction varies with the angle of incidence. The first is called the *ordinary,* the second the *extraordinary* ray. If the crystal be rotated, keeping the angle of incidence constant, the ordinary image will remain at rest, while the extraordinary image rotates about it in such a way that the line joining the two images lies in a *principal section,* a plane including the normal to the surface and an optic axis. If the ordinary and extraordinary rays *o* and *e* pass through a second crystal each ray generally divides in two, the rays *oo′* and *oe′* and the rays *eo′* and *ee′* (Fig. 586), the line joining each pair lying in a principal section of the second crystal. This gives rise to four images of the source, which are of equal intensity when the principal sections of the two crystals are at an angle of 45° with each other. If this angle be changed one pair of images will increase in intensity and the other diminish. When the principal sections are parallel only the rays *oo′* and *ee′* emerge; when they are at right angles, only the rays *oe′* and *eo′*. From such experiments Huyghens recognized the fact that light which has passed through Iceland spar, quartz, and other doubly refracting crystals does not possess properties which are alike in all azimuths around the direction of propagation. Newton, in

order to explain this, supposed the light corpuscles to be endowed with polarity of some sort—hence the name polarized light.

753. Direction of Vibration.—Fresnel explained the phenomenon of double refraction as a result of the transverse vibration of light waves. If the vibrations were longitudinal, it is impossible to conceive how they could be affected by rotation of the crystal in a plane at right angles to the direction of propagation. Transverse vibrations in a cord may be said to be polarized. Such vibrations would be freely transmitted through a slot parallel to the direction of vibration, but not through one at right angles to this direction. Longitudinal vibrations in a cord could be freely transmitted through a slot, regardless of its position. Fresnel assumed that in ordinary white light successive waves reaching a given point of space vibrate in different planes at random, so that, although each individual wave is vibrating transversely in a definite plane, and is, therefore, polarized, this direction changes so rapidly that the eye cannot take account of it and no polarization effects are observed. In passing through a doubly-refracting crystal vibrations in one direction travel with a different velocity from those in another direction, on account of the difference of the physical properties of the crystal in these directions, hence double refraction results. The displacement in each wave is in general resolved into two components, unless the light is travelling parallel to the axis. In that case it is unmodified, as the velocity of propagation is independent of the azimuth. In the ordinary ray, which travels in all directions with the same velocity, the vibrations must be at right angles to the optic axis. So long as this is the case the displacements will take place under the same conditions in every azimuth and the velocity be unchanged. In the extraordinary ray the vibrations must be in a principal section. This accounts for the fact that the ordinary and the extraordinary image are always in a line parallel to a principal section.

754. The Wave Surfaces.—From the experiments described above it may be seen that a wave of ordinary light on entering a doubly-refracting crystal is divided into two waves, one of which, *o*, has the same velocity in all directions in the crystal. The other wave *e* has a velocity which varies in different directions, and is the same as that of the ordinary wave only when both travel in the

direction of the optic axis. Huyghens showed that these facts are consistent with the existence of a double wave surface in the crystal, a sphere and an ellipsoid of revolution, which are tangent to each other at the two points where they intersect an optic axis. In one class of crystals, like Iceland spar, the sphere is inside the ellipsoid, and the ordinary wave is the more refracted (Fig. 587). In another class, represented by potassium sulphate or quartz,

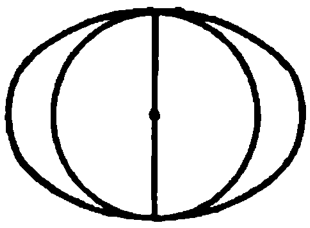


FIG. 587.

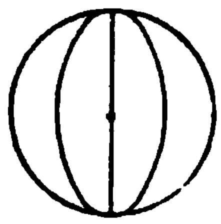


FIG. 588

the sphere encloses the ellipsoid, and the ordinary ray is less refracted (Fig. 588). The first are called negative and the second positive crystals. In the case of quartz the two wave surfaces do not touch where they intersect the optic axis, and there is double refraction of another kind in the direction of the axis (§771).

In crystals in which the physical properties are different along three axes at right angles to each other, such as sugar and topaz, which likewise show double refraction, there are two axes of no double refraction; hence such crystals are said to be *biaxial*, as contrasted with the class described above, which are said to be *uniaxial*. Both rays in biaxial crystals are extraordinary, that is to say, do not conform to the ordinary laws of refraction.

755. Double Refraction by Tourmaline.—Tourmaline is a semi-transparent hexagonal crystal. If light falls on a crystal, part is transmitted. If this falls on a second plate with its axis parallel to that of the first (Fig. 589), some of the light gets through; but if the second crystal is rotated about the line joining the two, less light gets through, and when its axis is at right angles to that of the first none is transmitted. Evidently the waves have had their mode of vibration so changed by passage through the first plate that they cannot pass through the second unless the principal sections of the two

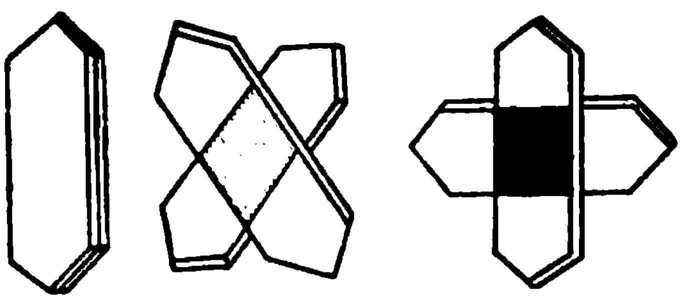


FIG. 589.

are parallel. If the light first passes through Iceland spar it is found that the extraordinary ray alone will pass through tourmaline if the principal sections of the two crystals are parallel, the ordinary ray alone if they are at right angles. It follows that light is doubly refracted by tourmaline, but that the ordinary ray is totally absorbed.

As a remarkable example of Kirchhoff's law (§736), it may be mentioned that if tourmaline is raised to a high temperature it emits polarised radiation. If this falls on a second crystal parallel to the first it is absorbed, showing that it corresponds to the ordinary ray. The mode of vibration which is absorbed corresponds to that which is emitted.

756. Polarization by Reflection.—About 1808 Malus discovered that light reflected from glass at a definite angle acquires properties similar to that of light transmitted through tourmaline or Iceland spar. When light is polarized by reflection from a mirror *A* (Fig. 590*a*) a large fraction is reflected from another mirror *B* if the two planes of incidence coincide. If the planes of incidence are at right angles (Fig. 590*b*) very little is reflected. If the light reflected from a glass plate is examined through a crystal of Iceland spar, the ordinary ray alone is transmitted when the plane of reflection coincides with a principal section, the extraordinary ray alone when the two are at right angles. In intermediate positions portions of both rays are transmitted.

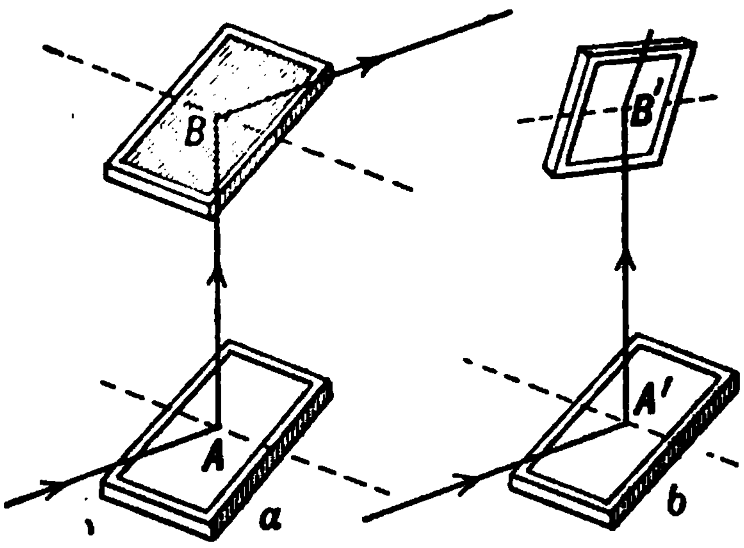


FIG. 590.

Similarly light reflected from glass is not transmitted through tourmaline if the plane of reflection is parallel to the optical axis of the crystal.

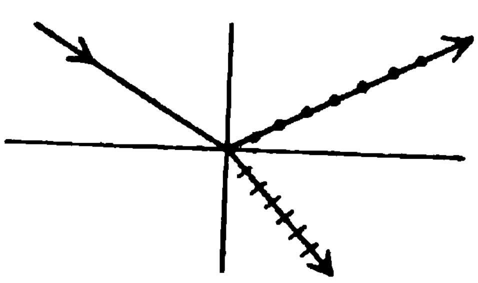


FIG. 591.

The simplest explanation of these effects seems to be that when light strikes a reflecting surface there is a partial resolution into components respectively in and at right angles to the plane of incidence. The vibrations parallel to the surface are most freely reflected,

while the others strike down into the surface and are transmitted or absorbed (Fig. 591). If polarized light is incident at the angle of maximum polarization (§759) on a piece of glass a large proportion will be reflected when its vibrations are parallel to the surface; if the vibrations are in the plane of incidence it will be refracted. In general both components are reflected and refracted, but the reflected light contains a larger proportion of waves vibrating perpendicularly to the plane of incidence, and the refracted light a larger proportion of the waves vibrating parallel to that plane.

It is clear that no interference effects can be produced between two vibrations in planes at right angles to each other.

This fact enabled Wiener to determine the direction of vibration in light polarized by reflection. A beam polarized by the mirror M fell at an angle of 45° on a thin transparent photographic film above a reflecting surface. He found that stationary waves (§700) were produced when the plane of incidence on the film coincided with the plane of reflection from the mirror (Fig. 592A), but this was not the case if the two planes were at right angles to each other (Fig. 592B). From the figure it appears that in the first case the vibrations must have been parallel to the film, and therefore to the mirror, in order that the incident and reflected rays should be in a condition to interfere at P, while in the second case the vibrations must have been in the plane of incidence on the film, and therefore parallel to the mirror, in order that the vibrations should meet at P at right angles to each other. This demonstrates that the vibrations in light polarized by reflection are parallel to the mirror.

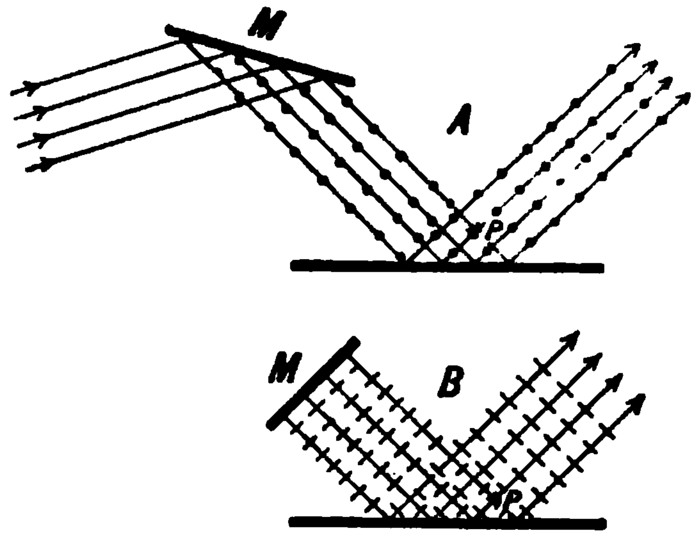


FIG. 592.

757. Plane Polarized and Ordinary Light.—The experimental evidence warrants the assumption that light waves are excited by the vibrations of the particles of material sources, these particles being probably ions or electrons within the molecules (§729); that these particles in general vibrate in different planes and directions, and that the vibrations of a given particle may constantly change in direction; that each vibrating particle sends out into the surrounding medium a series of waves vibrating in the same plane as the particle, so that ordinary white light

consists of a mixture of waves of many lengths, the resultant vibrations being in a plane at right angles to the direction of propagation, successive trains of waves having different planes of vibration; and that by double refraction or reflection we may sift out component vibrations in a given plane, and produce what is called polarized light. *When all the vibrations are in parallel planes the light is said to be plane polarized.* If such light is mixed with ordinary light, it is said to be *partially polarized.* If phase differences are introduced between two vibrations at right angles, the resultant displacement may be elliptical or circular (§243.) This gives rise to *elliptically* or *circularly* polarized light.

758. Plane of Polarization.—Before the direction of vibration in polarized light was known it became customary to speak of the "plane of polarization" of a polarized beam, rather than of the direction of vibration, and this plane was so defined that it coincides with the plane of incidence when the light is polarized by reflection. It follows that *the vibrations in a polarized beam are at right angles to the plane of polarization.*

759. Brewster's Law.—The light reflected from a surface is not in general completely polarized, that is, all its vibrations are not strictly in one plane. It is found, however, that for each reflecting substance there is a certain angle of incidence for which the polarization is a maximum. This is called the *polarising angle.* It was found by Fresnel that complete polarisation is given only by substances having an index of refraction equal to about 1.46. Brewster found that the polarizing angle is such that the reflected and the refracted rays are at right angles to each other. Since $n=(\sin i)/(\sin r)$ and since, when $i=p$, the polarizing angle, $p+r=90°$,

$$n=(\sin p)/(\cos p)=\tan p$$

From this relation the polarizing angle p may be determined. This is known as Brewster's law.

When the angle of incidence is different from that defined by this relation, and even for that angle when the index of refraction differs appreciably from 1.46, a small part of the component at right angle to the plane of incidence is reflected, with a phase different from that of the other component, resulting in elliptically polarised light (§769).

760. Pile of Plates.—Since only a small fraction of the incident light is reflected from a transparent substance, even when the reflected light is completely polarized, that which is refracted will be only partially polarized; that is to say, along with light vibrat-

ing in the plane of incidence a considerable proportion of that vibrating at right angles to this plane will be transmitted. If it is subject to a second reflection the proportion of polarized light is increased. After passing through eight or ten plates the transmitted light is almost completely polarized. If a pile P of thin glass plates is built up as shown in Fig. 593, the beam R, the result of successive reflections, and the beam T, which is transmitted, are completely polarized in planes at right angles to each other. This is one of the simplest methods of securing polarized light.

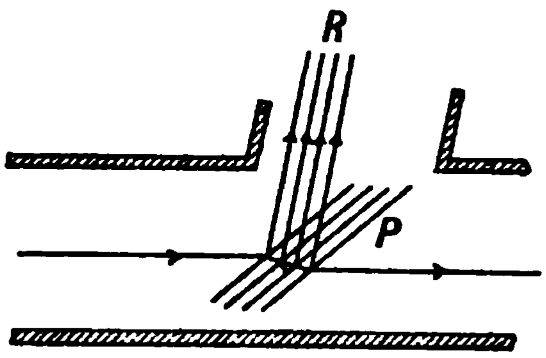


FIG. 593.

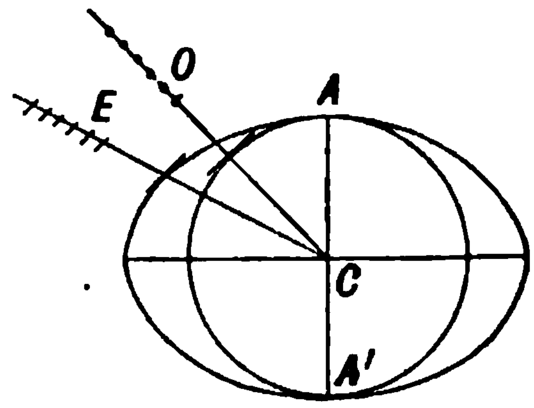


FIG. 594.

761. Wave Front Construction.—If C is a radiant point in a crystal of Iceland spar (Fig. 594) and if AA' is the optic axis passing through that point, two waves will diverge from C, one spherical and the other spheroidal. These waves will have the same velocity along AA', but in other directions the extraordinary wave will travel faster than the ordinary. The vibrations in each wave will be in the wave surface. The vibrations in the ordinary ray will everywhere be at right angles both to the optic axis and to the direction of propagation. In the extraordinary wave the vibrations are in general oblique both to the optic axis and to the direction of propagation of the disturbance. In this case we have an exception to the general rule that the wave normal indicates the direction of propagation.

By the application of Huyghens' principle the wave fronts in double refraction may easily be determined. Consider a plane wave AB incident on a crystal so cut that the optic axis is parallel to the surface and to the plane of incidence (Fig. 595). The two disturbances in the crystal will travel to O and E respectively while the wave travels from B to C in air. The tangent planes CO and CE are the two wave fronts. The disturbance at E is due to A, a point not on the normal to the extraordinary wave front passing through E. The wave velocity, or the velocity of the wave front, is proportional to the normal distance AN; the ray velocity, or actual velocity of the disturbance, is proportional to AE.

When the axis is parallel to the surface, but at right angles to the plane of incidence, the wave front is found as shown in Fig. 596. In this case, the extraordinary wave also has a circular section. Only in this plane of incidence is the ratio $(\sin i)/(\sin r)$ constant for the extraordinary ray,

and this ratio is called n_e, the extraordinary index of refraction. The value of the ratio V/V_e varies with the direction in every other plane of incidence, and hence cannot properly be called the index of refraction.

The general case, where the axis is at any angle with the refracting surface and the plane of incidence, is shown in Fig. 597.

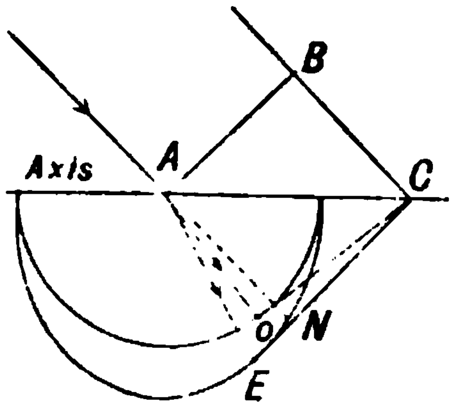


Fig. 595.

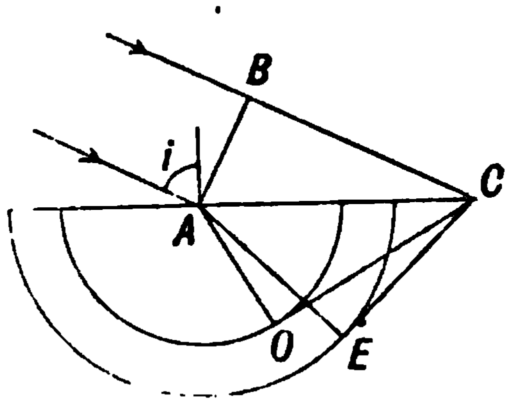


Fig. 596.

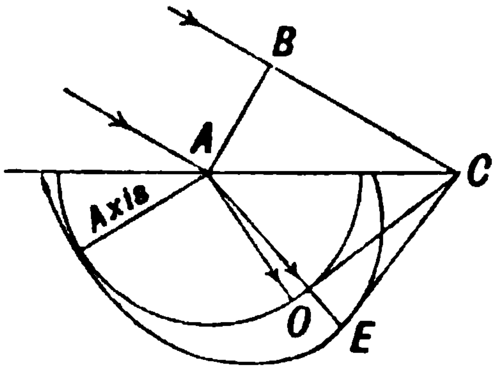


Fig. 597.

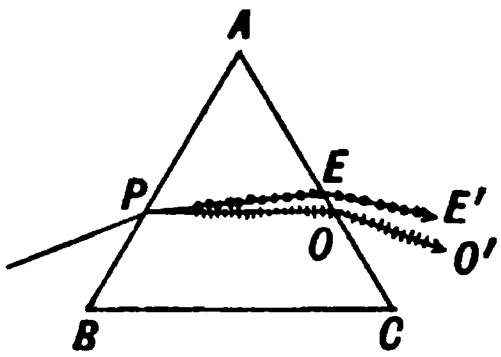


Fig. 598.

762. Uniaxial Prisms.—When light is incident on a doubly-refracting prism with its axis parallel to the refracting edge (Fig. 598), the ordinary and the extraordinary rays will be separated, and the angular divergence will persist after emergence. Two spectra will be formed, with light polarized in opposite planes. The ordinary spectrum will be less deviated than the extraordinary by a quartz prism and more by a calcite prism. When the optic axis is parallel to the refracting edge of the prism the two indices of refraction may be determined from the re'ations

$$n_o = \frac{\sin \frac{1}{2}(A + D_o)}{\sin \frac{1}{2} A} \text{ and } n_e = \frac{\sin \frac{1}{2}(A + D_e)}{\sin \frac{1}{2} A}$$

Some values of the indices of refraction for sodium light are given below:

Positive Crystals:	n_o	n_e
Quartz	1.5442	1.5533
Ice	1.3091	1.3104
Negative Crystals:		
Calcite (Iceland spar)	1.6584	1.4864
Beryll	1.5740	1.5674
Sodium nitrate	1.5874	1.5361

The difference between n_o and n_e is greater in the case of Iceland spar than in any other ordinary crystal.

763. Polarizing Prisms.—The two polarized rays produced by a doubly refracting crystal are not sufficiently separated to be conveniently used when a single beam is desired. The separation may be increased by using an ordinary triangular prism, but this introduces dispersion, so that other devices must be employed. The most common is the rhombohedral prism invented by Nicol, of Edinburgh, in 1828. In the principal section of a crystal of calcite (Fig. 599) the angles at B and D are 71°. The two end faces AB and CD are cut down to $A'B$ and $C'D$, so that these angles are reduced to 68°. The crystal is then sliced along $A'C'$ in a plane perpendicular to the ends and to the principal section. The two surfaces are polished and cemented together with Canada balsam, which has an index of refraction less than that of the calcite for the ordinary and greater for the extraordinary ray. If a ray of light r is incident in a direction parallel to the edge AD the ordinary ray will be totally reflected from the Canada balsam, while the greater portion of the extraordinary ray will be transmitted. The reduction of the angles at A and D is for the purpose of securing the proper angle of incidence on the balsam to produce this effect.

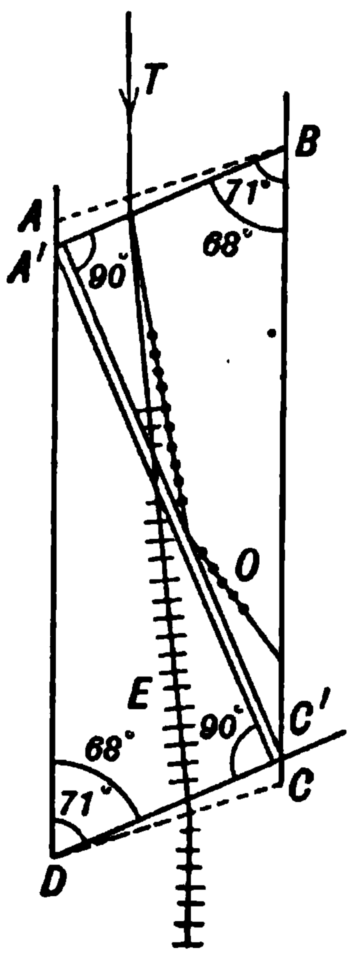


FIG. 599.

The Foucault prism resembles that of Nicol, but the total reflection is from an air film. This allows the prism to be made shorter, but there is a greater loss of light by reflection and a smaller field of view.

764. The Polariscope is an instrument for the study of the optical properties of substances with respect to polarized light. It consists of two Nicol prisms or piles of plates, one called the *polarizer*, to produce the polarized light, the other, the *analyzer*, which may be set with its principal section at any desired angle with that of the polarizer, to test the incident light with respect to the nature and direction of its polarization. If any doubly refracting substance is placed between the two its effects on the polarized light transmitted through it may be studied by the analyzer.

765. Resolution and Composition of Vibrations.—If the polarizer and analyzer are set with their principal sections parallel, light which has traversed the first will pass through the second without sensible loss. If their principal sections are at right angles to each other, or "set for extinction," no light will be transmitted through the analyzer. If the angle between the principal sections is α (Fig. 600), and if a is the amplitude of the light transmitted by the first Nicol, the amplitude of that transmitted through the second is $a \cos \alpha$, and its intensity is proportional to $a^2 \cos^2 \alpha$. The intensity of the totally reflected ordinary ray is $a^2 \sin^2 \alpha$. The sum of the two intensities is $a^2(\cos^2\alpha + \sin^2\alpha) = a^2$, which is equal to the intensity of the light incident on the analyzer. This simple law of resolution of vibrations into components by double refraction, giving determinate control of the intensity through a wide range, is made use of in several forms of photometer.

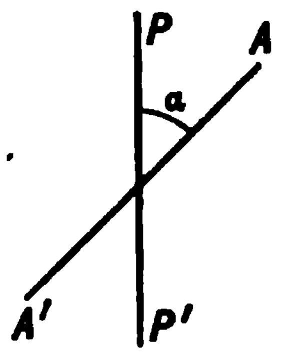


FIG. 600.

If the two Nicols are replaced by two crystals of calcite with their principal sections at an angle of α with each other, as in Huyghens' experiment (§ 752), an ordinary ray o and an extraordinary ray e of the same amplitude a are produced by the resolution of the vibrations along two directions in the first crystal. At incidence on the second crystal, the ordinary ray will be resolved into the components oo' and oe' of amplitudes $a \cos \alpha$ and $a \sin \alpha$ and the extraordinary ray into the components eo' and ee', of amplitudes $a \sin \alpha$ and $a \cos \alpha$. There will be, therefore, in general four rays, as found by Huyghens, which will be of equal intensity when $\alpha = 45°$. When the principal planes are at right angles, the incident ordinary ray goes through the second crystal as an extraordinary ray and the extraordinary as an ordinary ray, and there are only two images.

If the second crystal is replaced by a Nicol prism, with its principal section parallel to that of the first crystal, only the components oe' and ee' emerge, their vibrations being in the same plane, that of the principal section of the analyzer. If the two rays are superimposed on emergence, the intensity will depend not only on the amplitudes of the two components, but on the phase differences which have been introduced owing to the difference in velocity in the crystal of the two rays from

which these components are derived; in other words, there may be interference provided the light falling on the first crystal is plane polarized (see next section.)

766. Interference of Parallel Polarized Light.—If parallel plane polarized white light passes through a double refracting crystal of uniform thickness t and then through an analyzer, uniform colored effects are produced over the entire field, since some colors are reënforced and some weakened by interference. There is no real loss or gain for any color, for, as shown in §765, whatever energy is lost in the extraordinary ray is gained by the ordinary, and conversely, so that the ordinary light which is internally reflected in the prism is complementary to that passing through. When the principal sections of the crystal and the analyzer are either parallel or at right angles to each other no modification of the light is produced, the ordinary or the extraordinary ray alone getting through, so that there can be no interference. In all other positions of the analyzer there are varying proportions of white and colored light transmitted, the color effects being most pronounced when the principal sections are at an angle of 45° with each other.

The original beam of light falling on the crystal must be plane polarized. If ordinary light is used the succession of waves vibrating in different planes when resolved in the crystal will give rise to all possible distributions of amplitudes, so that all colors will be equally affected and the resultant effect will be white light.

767. Double Refraction due to Strain.—If a plate of glass or other isotropic substance is placed between a polarizer and an analyzer set for extinction no effect is produced. If the substance is then compressed or stretched some light will pass and interference effects similar to those described above will be produced. This shows that an isotropic substance becomes doubly refracting when subjected to unsymmetrical strain. This method offers a very delicate test of deviations from isotropy. Some liquids show the same characteristics in cases where the viscosity is so great or the stress so suddenly applied that a uniform hydrostatic pressure has not had time to become established throughout the substance. Imperfectly annealed glass exhibits double refraction. As shown by Tyndall, a bar of

glass set in longitudinal vibration restores the light through the crossed nicols, and a rotating mirror shows that the effect is set up periodically as the compression waves pass across the field.

Kerr found that a block of glass in a strong electrostatic field becomes doubly refracting like a uniaxial crystal with its axis parallel to the field.

768. Interference of Convergent or Divergent Light.—If a divergent or convergent pencil of polarised light falls on a doubly refracting crystal

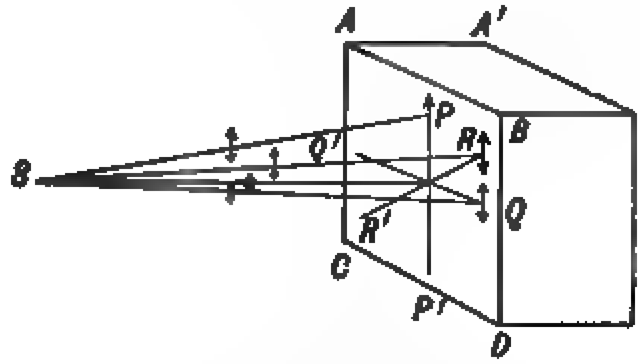


FIG. 601

different portions of the pencil will traverse the crystal at different angles, and therefore with different optical paths, hence the interference effects will not be uniform over the whole field. In general the effects are quite complex and cannot be discussed here, but the simple case of a uniaxial crystal cut perpendicularly to the optic axis may be considered as an illustration. Consider such a pencil diverging from S and falling normally on the face $ABCD$ of a doubly refracting crystal with its axis parallel to AA' (Fig. 601). The vibrations of the incident light may be supposed to be in a vertical plane, as indicated by the arrows. At P and Q the incident vibrations are respectively parallel and perpendicular to the principal sections PP' and QQ' and travel through without change. If an analyser is placed beyond the crystal and set for transmission or extinction of light transmitted by the polariser there will be a light cross or a dark cross on a screen beyond it corresponding to the crossed lines PP' and QQ'. The light incident at such a point as R, however, will be vibrating at an angle with the principal section RR', and will be resolved into two components. A relative difference of phase between them will exist at emergence, and interference effects will take place when they are re-resolved into the same plane by the analyser. The same difference of path will exist for all rays

FIG. 602.

incident at the same angle, that is, at all points equidistant from the normal from S to $ABCD$, hence colored rings similar to Newton's rings in appearance will be projected on a screen beyond the analyser. The "rings and brushes" due to a calcite plate are shown in Fig. 602. The brushes are dark, showing that the Nicols are crossed.

The interference effects due to crystals cut in other ways or to biaxial crystals are analogous to those described above, but more complex.

769. Circular and Elliptical Polarization.—Consider the state of the light originally plane polarised as it emerges from a doubly refracting crystal before it reaches the analyser. The ordinary and extraordinary rays start from the first surface in the same phase, but, as their velocities are different, one set of waves will fall behind the other. At different points within the crystal there will be two disturbances at right angles to each other and with phase differences depending upon the thickness of the medium traversed. The optical difference of path d at a distance t from the first surface is $[(V/V_e)-(V/V_o)]t$. At points where this difference is $n\lambda$ the light is plane polarised in a direction intermediate between the planes of vibration of the two components, the slope depending on their relative amplitudes, and being 45° if these are equal. If the difference of path is $(2n+1)/2.\lambda$ the light will likewise be plane polarised, but with a reversed direction of slope. If the difference of path is any odd multiple of a quarter of a wave-length the disturbance will be elliptical, or circular if the amplitudes are equal. For intermediate differences of path the disturbance will be elliptical, the axes of the ellipse being oblique with respect to the axis of the crystal. The successive stages at different distances from the first surface are shown in Fig. 603. On emergence from the crystal the disturbance will preserve the final form, and will be plane, elliptically, or circularly polarized according to the thickness of the crystal. If the waves are circularly polarised the disturbance travels through space like a point on a rotating screw. The polarization is said to be right-handed if the rotation is clockwise looking in the direction of propagation, or if the displacement resembles that of a right-handed screw, left-handed if the displacement is like that of a left-handed screw.

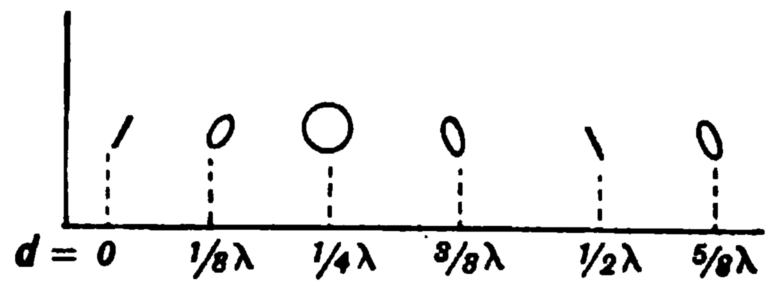


Fig. 603.

When light is totally reflected there is a phase difference between the vibrations respectively in and at right angles to the plane of incidence, so that this light is elliptically or circularly polarized. In ordinary reflection there is a slight elliptical polarization, which becomes very marked in the case of metallic reflection.

770. Production and Detection of Elliptically Polarized Light.—Circularly or elliptically polarized light cannot be detected by the unaided eye. If viewed through a Nicol prism no change in the intensity of circularly polarized light accompanies rotation of the prism, as a component of unchanging magnitude is transmitted. If the light is ellipitcally polarised there will be

variations of intensity as the prism is rotated, the intensity being greatest when the principal section of the prism is parallel to the major axis of the ellipse (component amplitude of greatest magnitude) and a minimum when it is parallel to the minor axis. If circularly polarized light passes through a crystal producing a relative retardation of an odd number of quarter wave-lengths of a particular color the additional retardation between the components will cause the emergent light to be plane polarized in an azimuth which may be found by the analyzing Nicol prism. Such a crystal is called a *quarter-wave plate.* These plates can readily be prepared from thin sheets of mica.

Another device for securing or testing circularly polarized light is *Fresnel's rhomb* (Fig. 604). A block of glass is cut with the angle at A equal to 54°, so that a pencil of light incident normally will be totally reflected at B and again at C, the angle of incidence being 54°. At each reflection at this particular angle a phase difference of an eighth of a period is introduced between the vibrations in and at right angles to the plane of incidence, and the emergent light is circularly polarized if the incident light is plane polarized at an angle of 45° with the plane of incidence. If this angle differs from 45° the amplitude of the two components will be different and the light will be elliptically polarized.

Fig. 604.

771. Rotation of the Plane of Polarization.—If two Nicol prisms are set for extinction and a crystal of quartz cut with the face on which the light falls at right angles to the axis, or a solution of sugar or tartaric acid, is placed between them, the light will be restored. On turning the analyzer through a given angle depending on the thickness of the crystal or the solution, the light will again be extinguished. This shows that the plane of polarization has been rotated through this angle. Substances producing this effect are said to be naturally optically active.

Some quartz crystals rotate the plane of polarization clockwise looking in the direction of propagation, and are called right-handed; others produce rotation in the opposite direction, and are called left-handed. These two classes of crystals can be distinguished by inspection on account of certain unsymmetrical facets which are differently placed in the two cases.

The rotation of the plane of polarization of light of the wave lengths corresponding to some Fraunhofer lines caused by a quartz plate of one mm. thickness is given below:

A	*B*	*C*	*D*	*F*	*G*	*K*
12.67°	15.75°	17.32°	21.70°	32.97°	42.60°	52.15°

As shown by these figures, the rotation varies very nearly inversely as the square of the wave-length.

Fused quartz shows no double refraction or rotation. These effects are evidently due rather to the crystalline arrangement of the molecules than to their individual structure.

If light passes through a quartz prism so cut that the light is transmitted in the direction of the optic axis it is found that there is a slight double refraction, so that spectral lines appear double. This shows that the two waves travel with slightly different velocities even along the optic axis; consequently the two wave surfaces cannot be tangent to each other (§754), but must be slightly separated. This is not generally true of uniaxial crystals, but only of those which rotate the plane of polarization. It is found that the two waves are circularly polarized in opposite directions, so that this is a case of circular double refraction. As first suggested by Fresnel, it appears that when light travels along the optic axis of quartz it is divided into two circularly polarized components, which travel with different velocities. These on emergence recombine to form plane polarized light but in a different plane. This offers a simple explanation of the rotation.

If each circular displacement r and l is resolved into two linear displacements x and y, it is seen that when the two velocities of propagation of the two circular components are equal (Fig. 605) the two x components at any point in the medium are equal and opposite, leaving the two y components

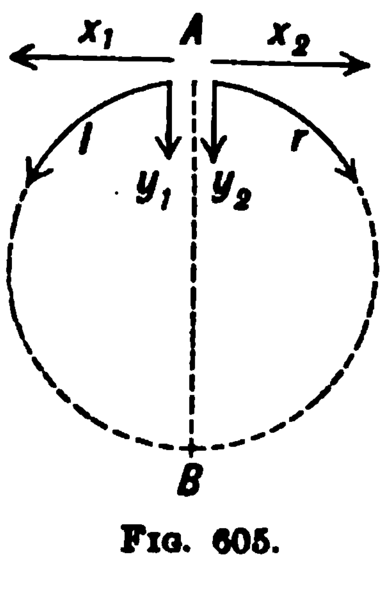


FIG. 605.

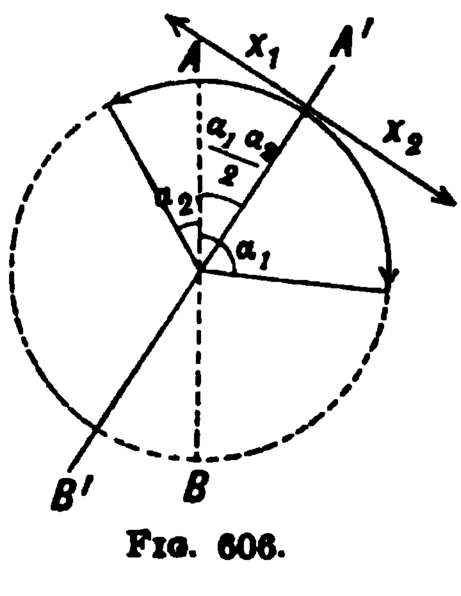


FIG. 606.

in the same direction to combine in a plane polarized beam, the vibrations of which are in the same direction as those of the original beam. But when the velocities of propagation are unequal (Fig. 606) the x components and the y components are respectively unequal. If, however, we refer displacements

to an axis of reference shifted through an angle $(\alpha_1 - \alpha_2)/2$ with the original direction of vibration it will be seen that with reference to this axis the x displacements will cancel each other. This line $A'B'$ then represents the final direction of vibration and the rotation is $(\alpha_1 - \alpha_2)/2$.

772. Rotation by Liquids. Saccharimetry.—A number of liquids, such as turpentine and the different sugars in solution, also cause rotation, that due to turpentine being left-handed, and that due to some sugars right-handed, of others left-handed. The vapors of such substances as turpentine also produce rotation. In such cases, as with quartz, there is circular double refraction, and the rotation is to be explained in the same way. In the case of liquids and vapors, however, the effect must be due to unsymmetrical structure of the molecule itself, as there is no crystalline structure, or if there is, the crystals are irregularly oriented. The amount of rotation varies inversely as the square of the wave-length, and is proportional to the thickness of the medium, and also to the concentration in the case of solutions.

Hence if α is the rotation of light of a definite wave-length in passing a distance of l (decimeters) through a substance of density ρ gm/cm^3 or of percentage concentration p, $\alpha=[\alpha]l\rho=[\alpha]lp/100$, where $[\alpha]$ is a constant for the substance called its *specific rotatory power*.

The rotatory power of the sugars is slightly affected by the presence of impurities. The percentage of sugar may be determined by measuring the rotation with a sensitive polariscope. This is called *saccharimetry*. Most sugars rotate to the right, but levulose rotates to the left. In some cases the specific rotatory power varies slightly with the concentration, and that of levulose is influenced by the temperature. The specific rotatory power for sodium light for some sugars at 20°C. is given below (from Landolt, Optical Rotation). The positive sign indicates right-handed, the negative left-handed rotation, while p is the concentration.

Sucrose (cane sugar)	$+\ 66.44° + 0.0087p$
Dextrose	$+\ 52.50° + 0.0188p$
Levulose	$-\ 88.13° - 0.2583p$
Lactose (milk sugar)	$+\ 52.53°$
Maltose (malt sugar)	$+140.4° - 0.0184p$

773. Rotation by Magnetic Field.—Faraday discovered that the plane of polarisation of light passing through a refractive substance in a magnetic field is rotated if the light travels parallel to the force lines. No effect is produced by a magnetic field on light waves in free space, and in general the effect increases with the refractive power of the substance, being especially marked in dense flint glass and carbon bisulphide and very feeble in the case of gases. The rotation is usually proportional to the field intensity and to the thickness of the medium. Some substances cause right-handed and others left-handed rotation. The effect varies with the wave-length. The rotation produced by 1 cm. thickness in a field of unit strength (Verdet's constant) is: For water, 0.0131°; carbon bisulphide, 0.0435°; dense flint glass, 0.06°. Enormous rotations are produced by thin films of iron or other magnetic material in a strong magnetic field.

In naturally active substances the direction of rotation is independent of the direction of propagation of the light, so that if a rotated beam is reflected its plane is turned back to the original position. In magnetically active substances the direction of rotation is reversed with reversal of the field, so that if the beam is reflected through the medium the rotation is doubled.

774. Kerr Effect.—When a beam of plane-polarised light is reflected from a metallic surface a relative phase difference is introduced between components respectively in and at right angles to the plane of incidence, so that the reflected light is elliptically polarized, unless the incident light vibrates parallel or at right angles to the plane of incidence. Kerr found that if the light is reflected from the polished pole of an electromagnet it becomes slightly elliptically polarized, even under the conditions just mentioned.

775. Zeeman Effect.—Zeeman placed a bunsen flame colored with sodium between the poles of a powerful electromagnet. When the light from the source traveled either parallel or at right angles to the direction of the field, he observed a broadening of the spectral lines when the field was established. H. A. Lorentz pointed out that such effects were in harmony with the *electron theory of radiation* proposed by him, and predicted that further investigation would show the radiation to be polarized by the field, either circularly or plane, according to the direction in which it was viewed. Zeeman found this to be the case. In the simplest cases, when the light is viewed normally to the field, each spectral line is split into triplets, the vibrations in the central and undisplaced component being parallel to the force lines, those of the lateral and displaced components at right angles to the force lines. When the source is viewed parallel to the force lines single lines become doublets, the components being circularly polarised in opposite directions, and displaced on each side of the mean position of the line. In some cases the effects are much more complex, a large number of components being produced from single lines, but the simple case described above is fully explained by Lorentz's theory which assumes that the light waves are disturbances caused by rotations of these electrons about the atoms of the source, and that the motion of the electrons is modified by the magnetic field.

DISPERSION AND SELECTIVE REFLECTION

776. Dispersion.—It was pointed out in §675 that dispersion due to refraction is irrational, that is, there is no simple relation between the deviation of lines in the spectrum produced by a prism of the substance and the wave-lengths, as there is in diffraction spectra. As a general rule the

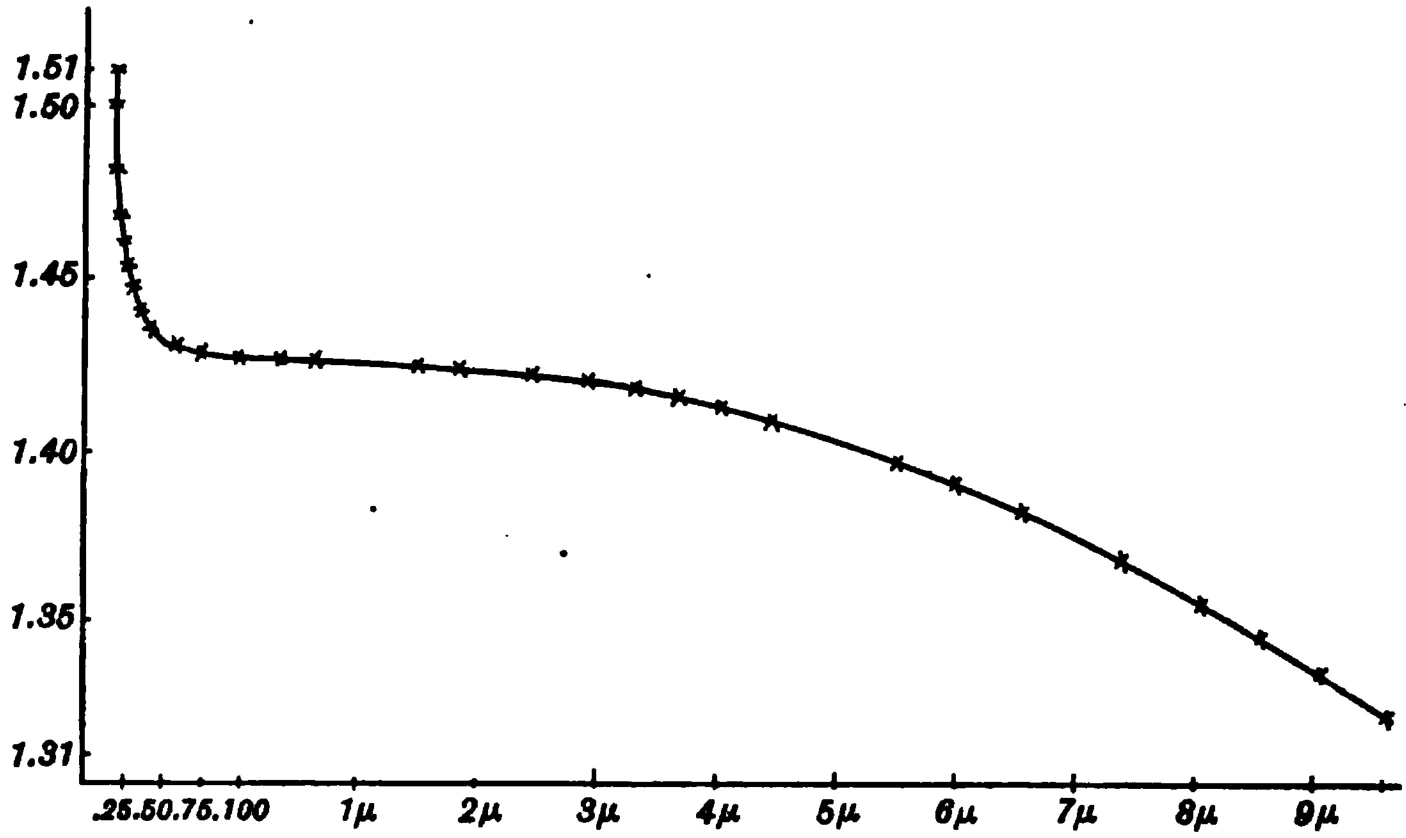


Fig. 607.

longer waves are less refracted than the shorter, and the dispersion steadily diminishes in the direction of the longer waves, so that the red end of the spectrum is "telescoped" as compared with the violet. Within the limits of the visible spectrum the relation between the index of refraction and the wave-length is closely expressed by the empirical relation (*Cauchy's* formula)

$$n = A + B/\lambda^2 + C/\lambda^4$$

where A, B, and C are constants varying with the substance. The dispersion curve of fluorite, showing the relation between index of refraction and wave-length, is shown in Fig. 607.

777. Anomalous Dispersion.—It is not always true that the deviation of waves by refraction increases as the waves become shorter. Iodine vapor transmits only the red and violet, and *the red is refracted more than the violet.* In the case of fuchsine, an aniline dye, blue and violet are less refracted than red, the green is absorbed, and the other colors occur in the usual order. Such anomalous dispersion is shown not only by a large number of substances such as the aniline dyes, but by the vapors of sodium and other metals, and, in fact, by almost every substance investigated in some part of its spectrum, visible or invisible. Anomalous

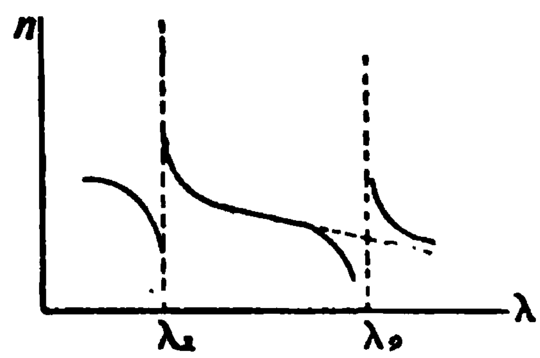


Fig. 608.

dispersion always occurs in the neighborhood of what appears to be a strong absorption band, which is, more properly, a region where the light is selectively reflected rather than transmitted or absorbed. The index of refraction is abnormally increased on one side of this band and diminished on the other, resulting in the reversal of the corresponding colors in a spectrum formed by a prism of the substance. The dispersion curve of a substance between two regions of such selective reflection is shown in Fig. 608. Between these regions the curve resembles the normal dispersion curve shown in Fig. 607.

Fig. 609.

The absorptive power of substances showing anomalous dispersion is usually so great that it is impossible to secure a prism of sufficient angle to give a spectrum long enough to clearly show the effect. The method of crossed prisms is well adapted for showing it. If light passes in succession through two prisms with their refracting edges at right angles to each other the resultant spectrum will usually be a line or smooth curve inclined in direction to the edges of both prisms. If, however, one of the prisms gives anomalous dispersion the resultant spectrum will be broken and irregular, as shown in Fig. 609, which illustrates the anomalous dispersion of sodium vapor in the neighborhood of the *D* lines.

778. Selective Reflection.—The color of natural objects is primarily due to selective absorption, the effective waves being those which escape absorption and become scattered. In the case of substances showing surface color, however, the effect is due to selective reflection, and the transmitted light is complementary to that reflected. This is the case with substances showing anomalous dispersion. The colors which are selectively reflected lie between the colors which are transmitted and anomalously dispersed. The so-called "absorption" bands so often referred to in this connection are thus seen to be largely due to lack of transmission because of reflection. As the reflecting power of the substance is abnormally high in such regions, they are said to show metallic reflection for the colors concerned. Recent investigations show that most substances exhibit anomalous dispersion in some region of their spectrum. For example, quartz, rock salt, and fluorite show anomalous dispersion and metallic reflection for certain very long waves. For radiation of wave-length 611,000 Ångström units (reflected from sylvite)

quartz has an index of refraction of 2.12, considerably greater than that of the shortest ultra-violet waves.

779. Theory of Anomalous Dispersion and Selective Reflection.—It is believed that these effects are due to resonance, the free periods of the vibrating parts of the molecules being the same as that of the waves selectively reflected. The vibrating element of the molecule is probably the electron. Selective reflection may be considered as the re-radiation of ether waves by the electrons, just as a tuning fork re-radiates sound waves after being excited by resonance. There is in such cases little "frictional" absorption of energy, which is completely transformed, to heat, not re-radiated. It may be shown from mechanical analogies and electrical theory that the rate of propagation of waves through a medium will be accelerated or retarded if the medium contains vibrating elements which have a free rate of vibration slightly greater or less than that of the waves.

A complete dispersion formula, taking account of regions having anomalous dispersion for wave-lengths λ_1 and λ_2 is

$$n^2 = A + \frac{B}{\lambda^2 - \lambda_1^2} + \frac{C}{\lambda^2 - \lambda_2^2}$$

where λ_1 and λ_2 are the lengths of the light waves having the same rate of vibration as the electrons of the substance. This gives a discontinuity in n, the refractive index, and anomalous dispersion for these wave-lengths.

The electron theory, first put on a definite basis by Zeeman's discovery, suggests an explanation of radiation and most of the optical properties of bodies.

References

Preston's *Theory of Light.*

Wood's *Physical Optics.*

Edser's *Light for Students.*

These three books give a more advanced treatment of the subject than the ordinary text-book, but contain much interesting material which can be understood by beginners.

Tait's *Light* also gives a somewhat advanced treatment.

Tyndall's *Light.*

Stokes' *Lectures on Light.*

Hasting's *Light.*

These three books give a very interesting and simple popular account of the subject.

S. P. Thompson's *Light, Visible and Invisible,* is an exceedingly interesting popular discussion of modern discoveries, not only in light, but in related phenomena, such as Röntgen rays and electric waves.

Le Conte's *Sight* treats of the eye and vision in a popular way.

Abney's *Color Measurement and Mixture* and *Color Vision* give an elementary but complete discussion of the subject.

Church's *Colour* is another interesting little book on the same subject.

Baly's *Spectroscopy* is an excellent presentation of spectroscopic methods and theory.

WATT's *Introduction to Spectrum Analysis* is a somewhat more popular book than the above, and gives tables of wave-lengths.

CLERKE's *Problems in Astrophysics* is a very interesting account of the uses of the spectroscope in astronomical work.

MICHELSON's *Light Waves and their Use* gives an account of interferometers and their applications.

DERR's *Photography* is an excellent elementary discussion of the photographic art.

The *Scientific Memoirs Series* contains the following reprints of important original papers, most of them presented in simple language: Prismatic and Diffraction Spectra, Fraunhofer; The Wave Theory of Light, Huyghens, Fresnel; Laws of Radiation and Absorption, Kirchhoff and Bunsen; The Effects of a Magnetic Field on Radiation, Faraday, Kerr, Zeeman.

Problems

1. A man is 5 feet 10 inches high. What is the shortest vertical plane mirror in which he can see his full-length image? *Ans.* 35 in.

Reflection.

2. Two plane mirrors are parallel to each other at a distance of 30 cm. Find the distance from each mirror of the three nearest images in each of an object between them and 10 cm. from one. *Ans.* 10, 50, 70; 20, 40, 80.

3. A beam of light is reflected from a plane mirror revolving clockwise about a vertical axis ten times per second, falls on a neighboring mirror revolving anticlockwise fifteen times per second, and then on a wall 10 meters away. What speed does the spot of light cross the wall? *Ans.* 3141.6 *m/sec.* anticlockwise.

A meter rod lies along the axis of a concave mirror of 20 cm. focal length, one end in contact with the mirror. Describe the images formed, and calculate the position of the first, fifth, tenth, twentieth, fortieth, and one-hundredth cm. marks, and the length of each division at these points (assuming the rod to be 2 cm. wide).
Ans. Virtual distances, 1.05, 6.67, 20, ∞ ; real, 40, 25. Lengths, 2.1 2.67, 4, ∞ ; 2, 0.5.

5. Prove by graphical construction the statements made in §666 concerning ellipsoidal, hyperboloidal, and paraboloidal mirrors.

6. Show by diagrams the successive shapes of the wave reflected from a hemispherical concave mirror as it passes from the mirror to a point beyond the focal cusp. (This surface must everywhere be normal to the "rays" which it cuts.)

7. A convex mirror has a focal length of 25 cm. Calculate the position and the height of the image of an object 10 cm. high and 15 cm. in front of the mirror. *Ans.* — 9.4, 6.3.

8. A paper square with sides two cm. in length lies in and parallel to the axis of the above mirror at a distance of 40 cm. Describe the shape of

the image, and calculate the lengths of its sides and the angles between them. *Ans.* Quadrilateral; sides normal to axis, 0.765, 0.744. Distance between them, 0.364; 91° 39′, 88° 21′

9. The sun has an angular magnitude of 32′. What is the size of the solar image formed by a concave mirror of 50 ft. focal length? *Ans.* 5.58 in.

Refraction. 10. A layer of ether ($n = 1.36$) 2 cm. deep floats on water ($n = 1.33$) 3 cm. deep. What is the apparent distance of the bottom of the vessel below the surface? *Ans.* 3.72.

11. An object is viewed through a cube of glass ($n = 1.55$) 10 cm. thick, in a direction at an angle of 60° with the normal to the glass surface. What is the lateral displacement of the image? *Ans.* 5.29 cm.

Lenses. 12. A convex lens 25 cm. from a candle-flame 5 cm. high forms an image of the latter on a screen. When the lens is moved 25 cm. further from the candle an image is again formed on the screen. Calculate the focal length of the lens, the distance of the screen from the candle, and the size of the two images.

Ans. 16.67; 75; 10, 2.5.

13. Show by graphical construction whether it is possible to construct a single thick double convex lens which will give a real erect image; and another which will give an inverted virtual image.

14. A candle flame 100 cm. from a convex lens of focal length 90 cm. is displaced 2 cm. away from the lens at the rate of 1 cm. per second. What is the displacement and the average velocity of its image?

Ans. 135 cm. toward lens; 67.5 $\frac{\text{cm.}}{\text{sec.}}$.

15. A convex lens ($n = 1.54$) has a focal length of 40 cm. in air. What is the focal length in water ($n = 1.33$)? *Ans.* 136.8.

16. The images of objects seen through a spherical flask or cylindrical glass of uniform thickness are of diminished size. Explain.

17. Two convex lenses of focal lengths 20 and 30 cm. are 10 cm. apart. Calculate the position and length of the image of an object 2 cm. long 100 cm. in front of the first lens. (Consider the image due to the first lens to be the object for the second.)

Ans. 10 cm. beyond second lens; length 0.33.

18. Replace the first lens in the above problem by a concave lens of the same focal length and determine the position and magnitude of the image.

Ans. 243 cm. to left of second lens; 3.04.

19. When focused on a star, the distance of the eye-piece of a telescope from the object lens is 50 cm. To see a certain terrestrial object clearly the eye-piece must be drawn out 0.2 cm. What is the distance of the object from the observer. *Ans.* 125.5 m.

20. In the above example, if the eye-piece has a focal length of 1 cm., and if the object referred to is a tree 10 feet high, what is the size of the image formed by the object lens? What is the angular magnitude of the image formed by the eye-piece? *Ans.* 1.21 cm.; 62° 21′.

21. A double convex lens with faces having a radius of curvature of 30 cm.

gives a real image at a distance of 60 cm. of an object 40 cm away. What is its focal length? Its index of refraction? *Ans.* 24; 1.625.

22. An achromatic lens is to be made of a combination of a crown glass double convex lens ($n_D = 1.51$, $n_F = 1.52$) and a plano-concave flint glass lens ($n_D = 1.64$, $n_F = 1.66$), the adjacent faces to fit together and the focal length to be 50 cm. Calculate the radii of curvature of the faces. *Ans.* $r_1 = r_2 = -r_3 = 19$; $r_4 = \infty$.

Photometry. 23. A candle is placed 10 cm. in front of a concave mirror of 20 cm. focal length (assumed to be a perfect reflector). What is the illumination on a screen 100 cm. from the candle along the mirror axis, as compared with I, that due to the candle alone? *Ans.* 3.31 I.

24. Solve the above problem after substituting a convex mirror of the same focal length for the concave mirror. *Ans.* 1.33 I.

25. Two sources have candle power 16 and 97 respectively. At what point between them must a screen be placed to be equally illuminated by the two? *Ans.* 0.288 d from fainter source.

26. Foucault in his arrangement for measuring the velocity of light (§644) placed the lens between the source and the revolving mirror. Show that with this arrangement—(*a*) The stationary mirror must be concave, with center of curvature at the axis of the revolving mirror; (*b*) that the stationary mirror cannot be placed far away from the revolving mirror unless its aperture is correspondingly enlarged, if the reflected beam is to have sufficient intensity. Show that Michelson's arrangement obviates these disadvantages.

Dispersion. 27. A 60° prism has an index of refraction of 1.62 for the D lines and 1.63 for the F line. If white light is incident at an angle of 45 degrees, what are the respective angles of emergence for these two colors? *Ans.* 65° 19′.8; 66° 40′.8.

28. In the above case, what is the angle of minimum deviation for each color? If the spectrometer telescope has a focal length of 30 cm., what is the length of the spectrum between D and F when the prism is set for minimum deviation for the D lines?

Ans. *D*, 48° 11′.2; *F*, 49° 10′.4; .71 cm.

Total Reflection. 29. Looking down into a cylindrical drinking glass partly filled with water, one cannot see external objects through the sides of the glass, but if a finger is firmly pressed against the side of the glass it can be seen from above. Explain.

30. Light incident internally on the surface of a glass prism at an angle of 56° is totally reflected from a drop of liquid in contact with the glass. If the index of refraction of the latter is 1.62 for sodium light, what is the index of refraction of the liquid? *Ans.* 1.343.

Interference. 31. In a system of Newton's rings due to a convex lens resting on a plane surface the 25th ring is 1 cm. from the center, when sodium light is used. What is the thickness of the air film at that point, and what is the radius of curvature of the lens?

Ans. 0.00751 mm.; 6.67 meters.

32. If the air film is replaced by water in the above example, what will be the distance of the 25th ring from the center? *Ans.* 0.97 cm.

33. Light from a narrow slit passes through two parallel slits 0.2 mm. apart. The interference bands on a screen 100 cm. away are 2.95 mm. apart. What is the wave-length of the light? *Ans.* 0.00059 mm.

34. The angles of a Fresnel biprism are 10′ and the index of refraction 1.62. What is the distance between the two images of a slit 20 cm. from the prism? What is the width of the interference bands of sodium light formed on a screen 50 cm. beyond the prism? What is their width if light of the wave-length of the F line is used? *Ans.* 0.724 mm.; 0.57 mm.; 0.47 mm.

35. A film of glass of index of refraction 1.54 is introduced in one of the interfering beams of a Michelson interferometer, and causes a displacement of 20 fringes of sodium light across the field. What is the thickness of the film? *Ans.* 0.0218 mm.

36. The D lines in the spectrum of the second order formed by a Rowland concave grating of 15 feet radius of curvature are 315 cm. from the slit. What is the distance between rulings? *Ans.* 0.00171 mm.

Diffraction. 37. The central maximum of the diffraction bands of sodium light produced by a narrow slit on a screen at a distance of 100 cm. is 2 mm. wide. How wide are the other maxima and the slit? *Ans.* 1 mm.; 0.589 mm.

38 Explain the diffraction bands in the shadow of a needle or wire (Fig. 558).

39. Describe and explain the appearance of the filament of a distant electric light seen through very small pinholes of different sizes.

40. Two narrow slits 0.1 mm. apart are illuminated by sodium light. What must be the diameter of a lens 5 meters away to clearly resolve the images of the two slits? *Ans.* 2.95 cm.

41. In the above case, at what distance will the same lens clearly resolve the images of the slits if they are illuminated by light of wave-length corresponding to that of the F line? *Ans.* 6.05 m.

Polarization. 42. Plane polarized light falls normally on a plate of quartz with faces parallel to the axis. If the vibrations of the incident light are at an angle of 30° with the principal plane, calculate the relative intensities of the transmitted ordinary and extraordinary rays. *Ans.* 0.25, 0.75.

43. In the above case, if the crystal is 1 mm. thick, what is the difference of phase upon emergence of the ordinary and extraordinary rays of sodium light (§762). *Ans.* 15.45 λ.

44. A crystal of Iceland spar cut with faces parallel to the axis is 2 cm. thick. How far below the upper surface are the ordinary and extraordinary images of a pencil mark on the lower face? *Ans.* 1.206, 1 346.

45. Through how many degrees will a column 20 cm. long of a 10 per cent. solution of cane sugar rotate the plane of polarization of sodium light? *Ans.* 132°.88.

Spectrum. 46. On mapping the spectral intensity curve of an incandescent source it is found that the maximum intensity

is at a wave-length 12,000 Ångström units. What is the temperature of the source? *Ans.* 2399° abs.

47. The displacement of the F line of hydrogen (wave-length 4861 Ångström units) in the spectrum of a star is .1 of a unit toward the violet. What are the direction of motion and the velocity of the star in the line of sight?

Ans. 6.2 $\frac{\text{k}}{\text{sec.}}$ toward earth.

Logarithms of Numbers from 1 to 1000.

No.	0	1	2	3	4	5	6	7	8	9
10	0000	0043	0086	0128	0170	0212	0253	0294	0334	0374
11	0414	0453	0492	0531	0569	0607	0645	0682	0719	0755
12	0792	0828	0864	0899	0934	0969	1004	1038	1072	1106
13	1139	1173	1206	1239	1271	1303	1335	1367	1399	1430
14	1461	1492	1523	1553	1584	1614	1644	1673	1703	1732
15	1761	1790	1818	1847	1875	1903	1931	1959	1987	2014
16	2041	2068	2095	2122	2148	2175	2201	2227	2253	2279
17	2304	2330	2355	2380	2405	2430	2455	2480	2504	2529
18	2553	2577	2601	2625	2648	2672	2695	2718	2742	2765
19	2788	2810	2833	2856	2878	2900	2923	2945	2967	2989
20	3010	3032	3054	3075	3096	3118	3139	3160	3181	3201
21	3222	3243	3263	3284	3304	3324	3345	3365	3385	3404
22	3424	3444	3464	3483	3502	3522	3541	3560	3579	3598
23	3617	3636	3655	3674	3692	3711	3729	3747	3766	3784
24	3802	3820	3838	3856	3874	3892	3909	3927	3945	3962
25	3979	3997	4014	4031	4048	4065	4082	4099	4116	4133
26	4150	4166	4183	4200	4216	4232	4249	4265	4281	4298
27	4314	4330	4346	4362	4378	4393	4409	4425	4440	4456
28	4472	4487	4502	4518	4533	4548	4564	4579	4594	4609
29	4624	4639	4654	4669	4683	4698	4713	4728	4742	4757
30	4771	4786	4800	4814	4829	4843	4857	4871	4886	4900
31	4914	4928	4942	4955	4969	4983	4997	5011	5024	5038
32	5051	5065	5079	5092	5105	5119	5132	5145	5159	5172
33	5185	5198	5211	5224	5237	5250	5263	5276	5289	5302
34	5315	5328	5340	5353	5366	5378	5391	5403	5416	5428
35	5441	5453	5465	5478	5490	5502	5515	5527	5539	5551
36	5563	5575	5587	5599	5611	5623	5635	5647	5658	5670
37	5682	5694	5705	5717	5729	5740	5752	5763	5775	5786
38	5798	5809	5821	5832	5843	5855	5866	5877	5888	5899
39	5911	5922	5933	5944	5955	5966	5977	5988	5999	6010
40	6021	6031	6042	6053	6064	6075	6085	6096	6107	6117
41	6128	6138	6149	6160	6170	6180	6191	6201	6212	6222
42	6232	6243	6253	6263	6274	6284	6294	6304	6314	6325
43	6335	6345	6355	6365	6375	6385	6395	6405	6415	6425
44	6435	6444	6454	6464	6474	6484	6493	6503	6513	6522
45	6532	6542	6551	6561	6571	6580	6590	6599	6609	6618
46	6628	6637	6646	6656	6665	6675	6684	6693	6702	6712
47	6721	6730	6739	6749	6758	6767	6776	6785	6794	6803
48	6812	6821	6830	6839	6848	6857	6866	6875	6884	6893
49	6902	6911	6920	6928	6937	6946	6955	6964	6972	6981
50	6990	6998	7007	7016	7024	7033	7042	7050	7059	7067
51	7076	7084	7093	7101	7110	7118	7126	7135	7143	7152
52	7160	7168	7177	7185	7193	7202	7210	7218	7226	7235
53	7243	7251	7259	7267	7275	7284	7292	7300	7308	7316
54	7324	7332	7340	7348	7356	7364	7372	7380	7388	7396
No.	0	1	2	3	4	5	6	7	8	9

Logarithms of Numbers from 1 to 1000.

No.	0	1	2	3	4	5	6	7	8	9
55	7404	7412	7419	7427	7435	7443	7451	7459	7466	7474
56	7482	7490	7497	7505	7513	7520	7528	7536	7543	7551
57	7559	7566	7574	7582	7589	7597	7604	7612	7619	7627
58	7634	7642	7649	7657	7664	7672	7679	7686	7694	7701
59	7709	7716	7723	7731	7738	7745	7752	7760	7767	7774
60	7782	7789	7796	7803	7810	7818	7825	7832	7839	7846
61	7853	7860	7868	7875	7882	7889	7896	7903	7910	7917
62	7924	7931	7938	7945	7952	7959	7966	7973	7980	7987
63	7993	8000	8007	8014	8021	8028	8035	8041	8048	8055
64	8062	8069	8075	8082	8089	8096	8102	8109	8116	8122
65	8129	8136	8142	8149	8156	8162	8169	8176	8182	8189
66	8195	8202	8209	8215	8222	8228	8235	8241	8248	8254
67	8261	8267	8274	8280	8287	8293	8299	8306	8312	8319
68	8325	8331	8338	8344	8351	8357	8363	8370	8376	8382
69	8388	8395	8401	8407	8414	8420	8426	8432	8439	8445
70	8451	8457	8463	8470	8476	8482	8488	8494	8500	8506
71	8513	8519	8525	8531	8537	8543	8549	8555	8561	8567
72	8573	8579	8585	8591	8597	8603	8609	8615	8621	8627
73	8633	8639	8645	8651	8657	8663	8669	8675	8681	8686
74	8692	8698	8704	8710	8716	8722	8727	8733	8739	8745
75	8751	8756	8762	8768	8774	8779	8785	8791	8797	8802
76	8808	8814	8820	8825	8831	8837	8842	8848	8854	8859
77	8865	8871	8876	8882	8887	8893	8899	8904	8910	8915
78	8921	8927	8932	8938	8943	8949	8954	8960	8965	8971
79	8976	8982	8987	8993	8998	9004	9009	9015	9020	9025
80	9031	9036	9042	9047	9053	9058	9063	9069	9074	9079
81	9085	9090	9096	9101	9106	9112	9117	9122	9128	9133
82	9138	9143	9149	9154	9159	9165	9170	9175	9180	9186
83	9191	9196	9201	9206	9212	9217	9222	9227	9232	9238
84	9243	9248	9253	9258	9263	9269	9274	9279	9284	9289
85	9294	9299	9304	9309	9315	9320	9325	9330	9335	9340
86	9345	9350	9355	9360	9365	9370	9375	9380	9385	9390
87	9395	9400	9405	9410	9415	9420	9425	9430	9435	9440
88	9445	9450	9455	9460	9465	9469	9474	9479	9484	9489
89	9494	9499	9504	9509	9513	9518	9523	9528	9533	9538
90	9542	9547	9552	9557	9562	9566	9571	9576	9581	9586
91	9590	9595	9600	9605	9609	9614	9619	9624	9628	9633
92	9638	9643	9647	9652	9657	9661	9666	9671	9675	9680
93	9685	9689	9694	9699	9703	9708	9713	9717	9722	9727
94	9731	9736	9741	9745	9750	9754	9759	9763	9768	9773
95	9777	9782	9786	9791	9795	9800	9805	9809	9814	9818
96	9823	9827	9832	9836	9841	9845	9850	9854	9859	9863
97	9868	9872	9877	9881	9886	9890	9894	9899	9903	9908
98	9912	9917	9921	9926	9930	9934	9939	9943	9948	9952
99	9956	9961	9965	9969	9974	9978	9983	9987	9991	9996
No.	0	1	2	3	4	5	6	7	8	9

Natural Sines and Cosines.

	Sine	D 1°		Cosine	D 1°
0	0.0000		90	1.0000	
1	0.0175	175	89	0.9998	02
2	0.0349	174	88	0.9994	04
3	0.0523	174	87	0.9986	08
4	0.0698	175	86	0.9976	10
5	0.0872	174	85	0.9962	14
6	0.1045	173	84	0.9945	17
7	0.1219	174	83	0.9925	20
8	0.1392	173	82	0.9903	22
9	0.1564	172	81	0.9877	26
10	0.1736	172	80	0.9848	29
11	0.1908	172	79	0.9816	32
12	0.2079	171	78	0.9781	35
13	0.2250	171	77	0.9744	37
14	0.2419	169	76	0.9703	41
15	0.2588	169	75	0.9659	44
16	0.2756	168	74	0.9613	46
17	0.2924	168	73	0.9563	50
18	0.3090	166	72	0.9511	52
19	0.3256	166	71	0.9455	56
20	0.3420	164	70	0.9397	58
21	0.3584	164	69	0.9336	61
22	0.3746	162	68	0.9272	64
23	0.3907	161	67	0.9205	67
24	0.4067	160	66	0.9135	70
25	0.4226	159	65	0.9063	72
26	0.4384	158	64	0.8988	75
27	0.4540	156	63	0.8910	78
28	0.4695	155	62	0.8829	81
29	0.4848	153	61	0.8746	83
30	0.5000	152	60	0.8660	86
31	0.5150	150	59	0.8572	88
32	0.5299	149	58	0.8480	92
33	0.5446	147	57	0.8387	93
34	0.5592	146	56	0.8290	97
35	0.5736	144	55	0.8192	98
36	0.5878	142	54	0.8090	102
37	0.6018	140	53	0.7986	104
38	0.6157	139	52	0.7880	106
39	0.6293	136	51	0.7771	109
40	0.6428	135	50	0.7660	111
41	9.6561	133	49	0.7547	113
42	0.6691	130	48	0.7431	116
43	0.6820	129	47	0.7314	117
44	0.6947	127	46	0.7193	121
45	0.7071	124	45°	0.7071	122
	Cosine	D 1°		Sine	D 1°

INDEX TO NAMES

Numbers Refer to Pages.

INDEX TO SUBJECTS

Numbers Refer to Pages.

3 2 5 —